UNIVERSITÉ DE MONTRÉAL

MESH-DERIVED IMAGE PARTITION FOR 3D-2D REGISTRATION IN IMAGE-GUIDED INTERVENTIONS

DAVID THIVIERGE-GAULIN
DÉPARTEMENT DE GÉNIE INFORMATIQUE ET GÉNIE LOGICIEL
ÉCOLE POLYTECHNIQUE DE MONTRÉAL

MÉMOIRE PRÉSENTÉ EN VUE DE L'OBTENTION
DU DIPLÔME DE MAÎTRISE ÈS SCIENCES APPLIQUÉES
(GÉNIE INFORMATIQUE)
AOÛT 2012

UNIVERSITÉ DE MONTRÉAL

ÉCOLE POLYTECHNIQUE DE MONTRÉAL

Ce mémoire intitulé :

MESH-DERIVED IMAGE PARTITION FOR 3D-2D REGISTRATION IN IMAGE-GUIDED INTERVENTIONS

présenté par : THIVIERGE-GAULIN David
en vue de l'obtention du diplôme de : Maîtrise ès Sciences Appliquées
a été dûment accepté par le jury d'examen constitué de :

M. GUIBAULT François, Ph.D., président
Mme CHERIET Farida, Ph.D., membre et directrice de recherche
Mme DUBOIS Josée, M.D., membre et codirectrice de recherche
M. KADOURY Samuel, Ph.D., membre

ACKNOWLEDGEMENTS

I would like to express my heartfelt gratitude to everyone that contributed to the realization of this project :

- First of all, to my thesis supervisor Dr. Farida Cheriet, who sparked my interest in the captivating field at the intersection of biomedical and computer engineering. I have learned much from our collaboration and am thankful for having such a great supervisor guide me through this academic adventure. Without your advice, patience and encouragements, this project would not have come to fruition.
- To my colleagues at the LIV4D, who have given me constant moral and academic support. You have made my stay at Polytechnique an enjoyable experience and I look forward to see you all soon again.
- To Dr. Josée Dubois and Dr. Sylvain Deschênes of CHU Sainte-Justine for their collaboration on the sclerotherapies for venous malformation project.
- At Siemens Corporation, Corporate Research and Technology, Dr. Atilla Kiraly and Christophe Chefd'Hotel for initiating the catheter ablation for atrial fibrillation project and Dr. Norbert Strobel from Siemens AX for providing the medical images. You support, guidance, numerous discussions and ideas were essential to the successful development of the project and are greatly appreciated. I have great memories of the internship at Siemens which was very gratifying and beneficial to my academic development. I also want to thank the other interns for making this experience unforgettable.
- To the organizations that have financed and participated to this project : NSERC's [1] MEDITIS program, the CHU Sainte-Justine, the LIV4D laboratory and Siemens Corporation, Corporate Research and Technology.
- To my family for an upbringing that galvanized my interest in science as well as continual moral support before and during this project.
- To the jury members, for taking time out of their busy schedule in order to evaluate this thesis.

1. Natural Sciences and Engineering Research Council of Canada

RÉSUMÉ

Les interventions guidées par images effectuées sous modalité 2D bénéficient de la superposition d'images 3D prises en stage préopératoire. La technologie nécessaire pour cette superposition est le recalage 3D-2D, qui consiste à trouver la position et l'orientation de l'image préopératoire 3D par rapport aux images intraopératoires 2D. Une intégration adéquate d'un algorithme de recalage à un processus chirurgical a le potentiel d'avoir un impact positif sur l'issue de la chirurgie et la durée de l'intervention. Cependant, beaucoup de chirurgies sont effectuées sans l'assistance du recalage, car aucune des solutions actuelles n'est applicable dans leur contexte clinique spécifique. Pour remédier à cette situation, cette thèse porte sur la recherche de solutions pratiques applicables à des interventions guidées par images spécifiques.

La première chirurgie étudiée est l'ablation par cathéter pour fibrillation atriale/auriculaire (AC pour FA) effectuée sous fluoroscopie rayons X, une procédure électrophysiologique traitant l'arythmie cardiaque. Dans cette chirurgie, une image volumétrique (soit résonance magnétique (RM) ou tomodensitométrie (TDM)) est prise avant l'opération pour définir l'anatomie de l'atrium gauche (AG) et des veines pulmonaires (VP)s. Un maillage segmenté de ce volume est ensuite utilisé pour offrir un support visuel intraopératoire lors du placement du cathéter d'ablation via sa superposition aux images fluoroscopiques. Cependant, les solutions de recalage actuelles sont trop lentes et requièrent des interventions manuelles, ce qui est problématique quand un recalage intraopératoire est nécessaire pour permettre de pallier aux mouvements du patient. Aussi, les solutions automatiques actuelles qui recalent les images 3D et 2D directement, sans passer par l'identification manuelle de points ficudiaux, ne sont pas assez précises pour être cliniquement utilisables. De plus, les solutions qui n'utilisent pas la cartographie électromagnétique ne fonctionnent pas avec les modalités RM/fluoroscopie rayons X. Ceci est un problème, car nous visons les interventions de AC pour FA qui utilisent la modalité RM sans la cartographie électromagnétique.

Il y a deux défis principaux pour arriver à une solution utile cliniquement. Premièrement, résoudre le difficile problème du recalage RM/fluoroscopie complexifié dans le cas de AC pour FA à cause de la correspondance partielle entre les modalités au niveau des VPs. Deuxièmement, de faire ce recalage assez rapidement pour permettre une mise à jour intraopératoire en temps réel dans les cas où le patient bouge pendant l'opération. Afin de remédier à cette situation, nous introduisons une nouvelle méthode de recalage basée sur la partition d'image dérivée d'un maillage (recalage PIDM). Cette méthode utilise les projections d'un maillage segmenté de la modalité 3D pour inférer une segmentation des images fluoroscopiques 2D. Ceci est beaucoup plus rapide que de faire des

projections volumétriques et, puisque le maillage peut être segmenté d'une image RM ou TDM sans distinction, la même procédure est valide pour les deux modalités. La justesse du recalage est évaluée par des mesures de similarité qui comparent les propriétés statistiques des zones segmentées et incorporent l'information de profondeur des maillages afin de tenir compte de la correspondance partielle au niveau des VPs.

Nous validons l'algorithme de recalage PIDM sur des interventions chirurgicales de AC pour FA provenant de 7 patients différents. Quatre mesures de similarité basées sur le principe de la partition à partir du maillage sont introduites et mises à l'épreuve sur 1400 cas biplans chacune. La précision, la portée et la robustesse de la solution sont évaluées en calculant la distribution de l'erreur (distance de projection) en fonction de la justesse de la pose initiale pour chacun des 5600 recalages. La précision est également évaluée de manière visuelle, en superposant les résultats du recalage et les valeurs-vérité sur les images fluoroscopiques. Pour donner une juste appréciation de la performance attendue de notre algorithme, les exemples visuels sont tirés de cas représentant l'erreur moyenne ainsi que d'un écart-type au-dessus et en dessous.

Afin d'évaluer l'extension du recalage PIDM à d'autres types de chirurgies, celui-ci est appliqué à des cas de sclérothérapie de malformation veineuse (SdMV). Ce type de chirurgie est particulièrement délicat à recaler car la malformation peut être présente sur toutes les parties du corps, ce qui offre peu de prévisibilité sur les propriétés des images médicales à recaler d'un patient à l'autre. De plus, cette chirurgie est effectuée en imagerie monoplan et les données ne sont pas accompagnées de méta-information permettant la calibration géométrique du système. Nous démontrons que le recalage PIDM est applicable aux cas de SdMV, mais doit être modifié pour être applicable à la grande variété de parties du corps où les malformations veineuses peuvent être présentes. Le protocole développé pour les chirurgies de AC pour FA peut être utilisé dans les cas où une embolisation ou une démarcation intérieure/extérieure d'une partie du corps est proéminente, mais il est nécessaire d'intégrer l'information de gradients dans les mesures de similarité pour recaler les organes où les os sont prédominants.

ABSTRACT

Image-guided interventions conducted under a 2D modality benefit from the overlay of relevant 3D information from the preoperative stage. The enabling technology for this overlay is 3D-2D registration: the process of finding the spatial pose of a 3D preoperative image in relation to 2D intraoperative images. The successful integration of a registration solution to a surgery has the potential for significant positive impact in terms of likelihood of treatment success and intervention duration. However, many surgeries are routinely done without the assistance of registration because no current solution is practical in their clinical context. In order to remedy these issues, we focus on producing practical, targeted registration solutions to assist image-guided interventions.

The first surgery we address is catheter ablation for atrial fibrillation (CA for AF), an electrophysiology procedure to treat heart arrhythmia conducted under X-ray fluoroscopy. In this surgery, a 3D image, either magnetic resonance (MR) or computed tomography (CT), is taken preoperatively to define the anatomy of the left atrium (LA) and pulmonary veins (PV)s. A mesh, segmented from the 3D image, is subsequently used to help positioning the ablation catheter via its overlay on the intraoperative fluoroscopic images. Current clinical registration solutions for CA for AF are slow and often require extensive manual manipulations such as the identification of fiducial points, which is problematic when intraoperative updates of the 3D image's pose are required because of patient movement. The automatic solutions are currently not precise enough to be used clinically. Also, the solutions which do not involve electroanatomic mapping are not suitable for MR/fluoroscopy registration. This is problematic since we target CA for AF interventions where the 3D modality is MR and electroanatomic mapping is not used.

There are two principal challenges to overcome in order to provide a clinically useful registration algorithm. First, solving the notoriously hard MR to X-ray fluoroscopy registration problem which is further complicated in cases of CA for AF because of the partial match between modalities at the level of the PVs. Second, solving the registration quickly enough to allow for intraoperative updates required due to the patient's movement. We introduce a new registration methodology based on mesh-derived image partition (MDIP) which uses projections of a mesh segmented from the 3D image in order to infer a segmentation of the 2D X-ray fluoroscopy images. This is orders of magnitude faster than producing volumetric projections and, since the mesh can be segmented from either MR or CT, the same procedure is valid for both modalities. The fitness of the registration is evaluated by custom-built similarity measures that compare the statistical properties of the segmented zones and incorporates mask-depth information to account for the partial match at the

level of the PVs.

We validate the MDIP algorithm on 7 cases of patients undergoing CA for AF surgery. Four MDIP-based similarity measures are introduced; each one is validated on 1400 biplane registrations. The precision, range, speed and robustness of the solution is assessed by calculating the distribution of projection distance error in function of the correctness of the initial pose for all 5600 biplane registrations. The precision is also evaluated visually by overlaying the ground-truths with results from the registration algorithm. To give a fair appraisal of the expected behavior, the examples are taken from cases exemplifying the average error measured as well as one standard deviation above and under.

The registration algorithm is also applied to cases of sclerotherapy for venous malformation (SfVM) in order to assess its portability to other type of surgeries. SfVM are especially challenging because the malformation can be present on any body part, which offers little predictability on the properties of the medical images from one patient to another. Our dataset is sampled from monoplane surgeries and did not come with metadata allowing a geometrical calibration of the system. We demonstrate that MDIP-based registration is applicable to cases of monoplane SfVM, but that modifications are required in order to account for the wide variety of body parts where VMs are common. The protocol developed for CA for AF surgeries can be used for embolizations or when the interior/exterior border of the organ is prominent, but gradient information has to be taken into account by the similarity measures in order to properly register cases where bones are predominant.

CONTENTS

LIST OF TABLES

LIST OF FIGURES

LIST OF ACRONYMS

#D	# Dimension, #-Dimensional
AF	Atrial Fibrillation
CA	Catheter Ablation
CT	X-ray Computed Tomography
DRR	Digitally Reconstructed Radiograph
EAM	Electroanatomic Mapping
GC	Gradient Correlation
IGI	Image Guided Interventions
IIBR	Intrinsic Image Based Registration
MDIP	Mesh-Derived Image Partition
MI	Mutual Information
mPD	Mean Projection Distance
mTRE	Mean Target Registration Error
MR	Magnetic Resonance
MRI	Magnetic Resonance Imaging, Magnetic Resonance Image
NCC	Normalized Cross-Correlation
NJH	Normalized Joint Histogram
PDF	Probability Density Function
PV	Pulmonary Vein
RMS	Root Mean Square

SfVM Sclerotherapies for Venous Malformation

SSD Sum of Squared Difference

TRE Target Registration Error

VM Venous Malformation

XMR X-ray/MR

CHAPTER 1 : INTRODUCTION

Image guided intervention (IGI) [1] is the use of medical images in order to assist surgical interventions. One of the earliest applications of IGI was reported in 1896 at McGill University in Montreal where a radiograph helped *planning* the removal of a bullet from the leg of a victim [1]. As medical-imaging systems improved, 2D imaging modalities matured from being solely planning and diagnosis tools to providing live intraoperative images during surgeries. Meanwhile, 3D imaging modalities such as magnetic resonance imaging (MRI) and X-ray computed tomography (CT) are now routinely used to *plan* the operations. The next logical step is to conduct surgeries with real-time intraoperative updates of the 3D modality. Unfortunately, the current state of medical imaging science is still far from achieving that goal. What is possible *today* is to overlay a static 3D image taken before the operation to the intraoperative 2D images. The gains from this augmented reality [2] is not only to display a third dimension but also to draw from the *complimentary* nature of the different modalities, none of which is superior to all others regarding all aspects [1]. This additional information improves the ability of the medical staff to make judicious decisions, potentially reducing operation time and cost as well as improving the average outcome for patients.

(a) X-ray fluoroscopy of the left atrium.

(b) MR image of the left atrium with aorta.

(c) X-ray fluoroscopy with superimposed preoperative MR image.

Figure 1 Example of a preoperative 3D image being overlaid to a 2D intraoperative image.

In order to reap the benefits from the overlaid 3D image, its position and orientation relative to the intraoperative 2D images — the *pose* of the patient/organ — needs to be correct. The process of finding that pose is called *3D-2D registration*. Unfortunately, even if operating rooms have all the necessary hardware, surgeons often conduct the operations without registration's assistance. It

1. Loosely associated appellations: image-guided surgery, image-guided therapy.
2. Also called: computer-assisted surgery, computer-aided surgery or computer-assisted intervention.

is possible have a medical specialist manually provide registration, to add expensive tracking hardware or add steps to the surgical procedures in order to help registration (such as the implantation of fiducial markers prior to surgery). However, medical professionals are rightfully reticent to change their operative protocol, buy additional hardware, add staff or procedure steps to their surgical workflow. This thesis focuses on *practical*, low-cost, automatic registration solutions that can be integrated without impeding the surgery while having the potential for significant positive impact on surgeries. In order to fully understand the challenges to providing such a solution, this chapter contains a brief introduction to the main target surgery: *catheter ablation for atrial fibrillation* (CA for AF: section 1.1). Even though registration is a mature research field, each surgery has unique challenges in terms of speed, precision requirements and modeling the statistical or stochastic link between multimodal medical images. This implies that a registration solution usually has to be fine-tuned to the target surgery.

In order to evaluate if the solution proposed is easily extendable to other type of surgeries, the same method as developed for CA for AF is also evaluated on *sclerotherapies for venous malformation* (SfVM), which are described in section 1.2.

1.1 Clinical Application: Catheter Ablation for Atrial Fibrillation

Atrial fibrillation (AF) is a heart rhythm disorder which affects more than two million people in the United-States [2]. It causes palpitations, shortness of breath and is a risk factor for strokes. AF is caused by abnormal electrical activity originating around the pulmonary veins (PV)s, which triggers abnormal beating (fibrillation) of the heart.

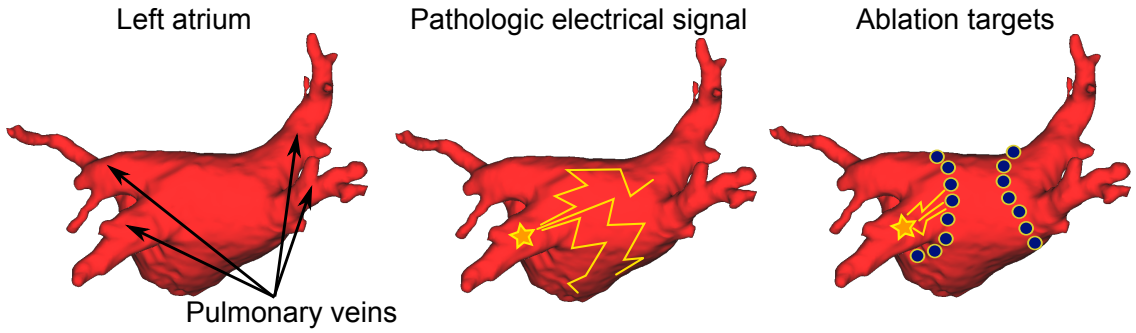
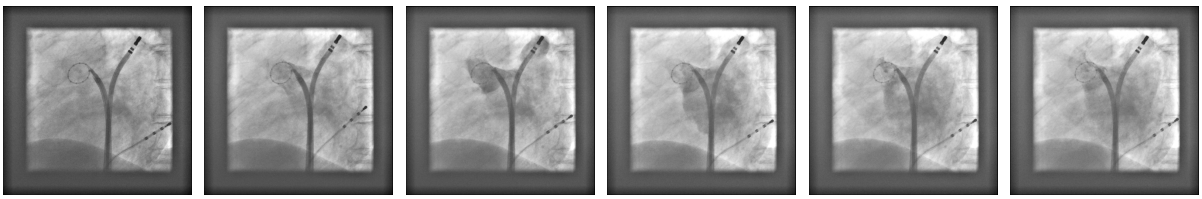


Figure 2 From left to right: LA with attached PVs, pathologic electrical current that triggers the arrhythmia, the ablation targets in CA for AF.

CA for AF is a surgery which aims at stopping the propagation of the problematic electricity by inflicting small lesions that act as a barrier to the pathological electric current. The lesions are administered via an ablation catheter dispensing radio-frequency energy which is inserted from

the groin or the side of the neck, then guided through a blood vessel into the heart. During the operation, a puncture is made on the wall separating the right and left atria to allow direct insertion inside the left atrium. The catheter is then moved to the target ablation sites, producing small cauterizations of 6 to 8 mm in diameter [2]. The ablations are distributed around the PVs which are the point of entry of oxygen-rich blood from the lungs into the left atrium (LA) (see figure 2). The cauterization caused by the ablation encircles the pulmonary veins, preventing the propagation of the problematic electrical current. The operation requires from 2 to 6 hours. The success rate of the operation is from 30% to 90%, depending on the severity of the disease. As a consequence of the operation, 1% to 2% of patients suffer life-threatening complications such as a puncture of the heart or stroke [2].

The operation can be done under monoplane or biplane fluoroscopy, with monoplane fluoroscopy being more frequent. Fluoroscopy is a 2D projective imaging modality which cannot distinguish the soft tissue of the LA and PVs without the injection of contrast agent [3].



(a) Frame 0 (b) Frame 5 (c) Frame 10 (d) Frame 19 (e) Frame 30 (f) Frame 35

Figure 3 Injection of contrast agent in the LA. Notice that in (a) — the frame without contrast agent — the LA and PVs are not visible. In the other frames, it is still difficult to identify anatomical regions: no single image offers a clear, comprehensive understanding. The surgeon has to watch the injection of contrast agent attentively and rely on his expertise to construct a mental representation of the location of important structures, which complexifies the procedure and increases the risk of complications [4].

Figure 3 reveals that the surgeon has to rely on a mental map of the LA during the catheter guidance and subsequent embolizations. This opens the door to costly, life-threatening errors. In order to reduce the likelihood of errors, computer-assisted image guidance is provided by overlaying a *mesh* extracted from a preoperative MR image to the intraoperative X-ray fluoroscopy images. This mesh allows a visual reference of the anatomical structures to be present after the contrast agent has washed away. This visual map reduces the number of injections of contrast agent necessary for guidance [3], as well as allowing a display of the past embolization sites as visual marks on the mesh. The visual marks enable a more accurate distribution of the ablation target sites, reducing the

3. There is a limit to the amount of contrast agent that can be safely injected in the patient.

potential for errors. The presence of the visual guidance system decreases the procedure's duration and likelihood of AF recurrence: 84% of patients who benefited from registration-based guidance had no recurrence of AF versus 64% in the control group (sample size of 50 patients) [3]. Another study done on 94 patients by Kistler et al. [5] reveals that the 3D-2D overlay of a CT image via registration for cases of CA for AF reduces the fluoroscopy time from 62 $\pm$ 26 min to 49 $\pm$ 27 min as well as a 32% arrhythmia recurrence compared to 52% when not using overlays. Since the intraoperative modality is X-ray fluoroscopy, a reduction in operation time also translates in a lower dose of radiation imposed on the patient.

Registration is the operation that positions a 3D preoperative image in relation to the 2D intraoperative images. It allows a 3D-2D overlay that can be used to assist catheter guidance and identification of ablation sites. This registration can be done manually, but is time consuming (typically takes a few minutes) and requires the training and presence of an additional skilled medical expert in the room. An automatic registration algorithm that systematically positions the LA at the proper position and orientation with an acceptable, reproducible and predictable error rate is therefore highly desirable. Registration algorithms achieve better results when using biplane images; with monoplane images it is likely that there will be large out-of-plane errors. To simulate two planes, the C-arm of monoplane systems can be rotated between subsequent image acquisitions. However, this reduces the registration accuracy due to non-simultaneous image acquisition which results in motion artifacts.

There are two main registration methodologies for cases of CA for AF that have clinical precedents: surface/landmark registration via an electroanatomic mapping (EAM) system and direct CT/fluoroscopy-based registration.

The surface/landmark registration method registers the preoperative MR or CT image to a surface obtained from an EAM system which gathers information from a specialized mapping catheter via coils positioned underneath the patient [6]. The registration is done using both manually identified fiducial points as well as 3D-3D surface-to-surface matching. Kistler et al. [5] report using this technique successfully on 47 patients (total 94 patients, 47 controls) using landmark points pairing of proximal or first order venous branches on PVs for a first crude registration, then refining with the surface-based method. The landmark registration error is 5.7 mm (measured on the PVs) and 2.4 $\pm$ 0.4 mm when using CT/EAM surfaces registration. Fahmy et al. [7] evaluated the precision of using different landmarks on 124 patients, and obtained the best results (an error of 5.6 $\pm$ 3.2 mm) when using the posterior points on the pulmonary veins. Surface registration is done by comparing an electroanatomic map of 40-50 points with the CT surface. Using the surface registration, they evaluate that the LA cavity is aligned with a precision of 2.17 $\pm$ 1.65 mm. However, the land-

mark error jumps from 5.6 ± 2.1 mm before the surface registration to 9.2 ± 2.1 mm after, casting doubt on the validity of the error assessment for the surface registration.

The CT/fluoroscopy-based 3D-2D registration operates directly on the medical images. It is considered a harder problem than the 3D-3D surface-based registration of the previous sections, but has the advantage of not requiring specialized equipment and surgical steps associated with EAM systems. Sra et al. [8] demonstrated the feasibility of a landmark registration based on the alignment of the coronary sinus catheter on the fluoroscopic images with the coronary sinus segmented from the CT. Tests done on 20 patients were only evaluated qualitatively and reveal that manual alignment after the automated registration was needed in 51% of the image sequences.

Knecht et al. [9] evaluate three different 3D-2D CT/fluoroscopy registrations on 56 patients. Two methods are similar to the methods of Sra et al. (based on manual alignment of landmarks or segmented heart contour) and are reported to have an accuracy of 0.3 ± 0.5 mm for the landmark method and 1.3 ± 0.6 mm for the method using the hear contour. The third method is based on the *automatic* registration of LA via registration of the spine from the segmented CT image and real-time fluoroscopic images. This method, highlighting the difficulties of automated systems, has an accuracy of 17.0 ± 9.0 mm. They report an average time of 7 ± 2 minutes for the algorithm to complete. They do the registration during the patient preparation and catheter insertion, and mention that as a result the operation is not lengthened by the registration. However, in the same article they mention that repeat registration was necessary in case where the patient moved during the procedure, indicating that long (minutes) delays for registration do impede some procedures.

Currently, clinically usable registration systems for CA for AF exist. The methods that use EAM mapping require manual steps and the acquisition of additional, expensive hardware, but have the advantage of being compatible with both CT and MR modalities. The direct CT/fluoroscopy-based systems usually require manual identification of fiducial points, or lack precision. The current systems are slow (∼7 minutes for registration), which can be a problem if registration has to be repeated during the operation because of patient movement. There is currently no system available that is compatible with the MR modality without using EAM mapping [10]. Using MR instead of CT is preferable since it reduces the harmful radiation dose imposed on patient (see section 1.3 for details on the health considerations of modalities which irradiate the patient). There is therefore a need for a new automatic registration solution that is fast, precise, and works for the MR/fluoroscopy modalities pairing.

1.2 Clinical Application: Sclerotherapies for Venous Malformation

Venous malformation (VM) is a congenial vascular abnormality. It is a spongy masses of veins of various size characterized by the absence or low-velocity venous flow which can be present anywhere inside the patient, but most commonly found in the head, neck, extremities and trunk [11]. Although asymptomatic in some patients, other experience thrombosis within the malformation, which results in the development of a firm mass, swelling and pain. Surgery is required when the VM inconveniences the patient or for aesthetic considerations. Since VMs are radioresistant and generally not treatable by excision, the surgery consists of injecting a sclerosing agent such as ethyl alcohol into the VM while the patient in under general anesthesia. In order to render the sclerosing agent X-ray visible, it is mixed with a contrast agent. The operation often requires repeated aggressive treatment because VMs have the tendency to recanalise after the sclerotherapy [12], and a wait period of 1 to 3 months must be observed between sclerotherapies. Before the operation, MR images are sampled to assess the extent of the VM and help plan the surgery. MR images can also be taken after the operation to assess if repeating surgery is warranted. However, because of transient inflammatory reaction to the sclerosing agent, a period of 6 months is advised before a MR assessment of the therapeutic response after sclerotherapy [11]. XMR rooms (**X**-ray/**MR**) allow the patient to be slid from the fluoroscopic table to the MRI during the operation, enabling intraoperative assessment of the sclerotherapy before inflammation occurs. As of now, the surgeon has to memorize the information gleaned from the MR image, and then use that information in the subsequent sclerotherapy. A properly registered overlay of the MR image to the fluoroscopic images would reduce mental load, allow for better precision and reduce the likelihood of error in embolization location.

1.3 Image-Guided Interventions Modalities

This section covers the modalities used in the surgeries for which we want to provide image-guidance: CA for AF and SfVM.

X-Ray

X-rays are an electromagnetic radiation of wavelength of 0.01 to 10 nanometers. When passing through biological tissue, X-rays are absorbed at a different rate depending on the properties of the type of tissue traversed. By measuring the attenuation of the X-ray energy, it is possible to create an image that depicts the internal properties of the tissue imaged. The absorption rate is high for bones, which makes it an ideal modality for orthopedic use, but is relatively weak for soft tissues [13].

The simplest way to form an X-ray image, **X-ray radiography**, is to place a patient between an X-ray source and a detection plane, which produces a static 2D image. **X-ray fluoroscopy** allows real-time images at 30 frames per second [14] by using a lower energy level. X-ray fluoroscopy is widely used as the interventional modality in cardiac surgeries. X-ray radiography and X-ray fluoroscopy have the limitation that the images produced are two-dimensional and projective by nature, resulting in an information loss that makes overlapping structures indistinguishable from one-another.

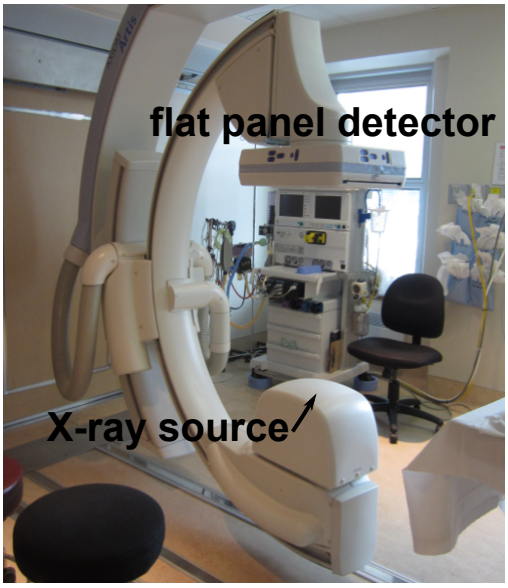


Figure 4 Monoplane X-ray fluoroscope.

X-ray computed tomography allows to produce *slices* of the patient's anatomy, circumventing the overlapping problem caused by the projective nature of X-ray radiography. The images are produced by a narrow energy beam that rotates around the patient, creating several 1D signals. A technique known as *back-projection* allows a high-resolution ($\approx$1 mm [13]) 2D image of a patient's slice to be reconstructed from 1D signals taken at different angles. An X-ray CT scan typically takes a few seconds. A helical CT can create a 3D volume of the patient by using the same method, but by continuously moving the patient perpendicular to the 2D space sampled by a rotating X-ray CT device. It is possible to get 3D images of the heart using a helical scan, but special precaution has to be taken in order to minimize movement artifacts due to the patient's movement, heartbeat and respiration. The images are taken while the patient is holding his breath, and is synchronized with an electrocardiogram in order to sample images from the same phase of the cardiac cycle.

However, X-ray-based image modalities irradiate the patient, which can have significant long-term negative health impact. This is especially true for CT scan, where the dose from a single abdominal scan delivers 1000 times more radiation than a single posterior-anterior chest radiogra-

phy. There is no large-scale epidemiological study that directly links CT exposure to cancer risk. However, a study by Brenner et al. [15] extrapolates known cancer risk for specific organs under different radiation doses to estimate the health impact of CT-scanning. Using that method, it is estimated that 0.4 to 2% of all cancers in the United-States are due to CT-irradiation. A single abdominal CT-scan of a patient in his thirties is expected to raise his likelihood of death from cancer of 0.05%. Since radiation is especially harmful to children, that risk raises to 0.09% if the patient is about 5 years old. Therefore, pediatric cases warrant extra care regarding irradiation exposure. Any reduction of irradiation exposure is a statistical improvement of the patient's future health prospect.

The visibility of the target organ (especially when it is mainly composed of soft-tissue) can be increased by the injection of a liquid which has a very high X-ray absorption rate, such as barium sulfate. This injection allows the organ to appear momentarily very clearly, but has to be used parsimoniously because it can harm the patient in large doses.

Magnetic Resonance

Magnetic resonance (MR) is a 3D imaging modality based on the interaction of a strong magnetic field, atomic nucleus and a radio-frequency pulse. The application of a strong magnetic field has for effect to align the spin of the patient's body protons in a parallel or anti-parallel direction. Then, a radio-frequency pulse is applied, causing the protons to precess in their alignment with the magnetic field. When the radio-frequency pulse is switched off, the protons' orientation gradually align back with the magnetic field, but at a different rate for protons belonging to different tissue types. It is the measurement of these different relaxation times that allows forming MR images (MRI) [4]. Since protons are in highest density in water and fat, MR is an ideal modality for soft-tissues; this contrasts with X-ray, which is a poor modality for those tissue types. This also makes multimodal registration arduous because the statistical properties of the intensities in CT/X-ray and MR images is inherently different and have a complex, hard to model relationship to each other. It also has the advantage over X-ray of being non-ionizing (there are no known health impact of MR imaging) while having a similarly high spatial resolution ($\approx$ 1 mm).

4. The acronym MRI is used to refer to either MR images or MR imaging. The meaning should be unambiguous from the context.

Figure 5 MRI scanner.

However, the slow image acquisition time make MR imaging susceptible to movement artifacts, which is especially problematic when the target organ is non-static and deformable. For example, in cardiac application, it is necessary to use an electrocardiogram to synchronize the MR acquisition to the cardiac cycle. MR is currently rarely used as the interventional modality. This is due to both the slower imaging time as well as the surgical tools being generally unfit for use in an MR scanner [14]. Although successful cardiac catheterization guided by MR has been reported [4], it is not the norm.

Because of the high-quality of image produced, MR is widely used as a diagnostic tool. For example, brain tumors have higher water concentration than the surrounding tissue. Parkinson's disease, Alzheimer's disease and hemorrhages result in deposits of iron which can be detected because of their interaction with water. In the same manner, MR is often used to diagnose liver diseases, musculoskeletal abnormalities as well as cardiovascular diseases [13].

The contrast agents used in MR imaging is typically gadolinium, manganese, europium or small iron particles possessing a high magnetic moment [13]. These agents reveal small regions or ease the segmentation of an organ, but the signal to noise ratio of MR images is often high enough to preclude their use.

1.4 Current Challenges and Research Goals

CA for AF and SfVM surgeries would benefit from a comprehensive, specifically-tuned registration solution. They are both commonly conducted without automatic registration and sometimes without any overlay of 3D preoperative images. Current clinically-used algorithms are slow, which can be problematic during surgeries. For CA for AF, there is no current solution for registration

without EAM with MR as the 3D modality. In order to bridge the gap between science and the operating room, we set the following goal:

– Provide *quick* and *automatic* registration that *does not impede the surgical procedure*. These goals can be achieved by *careful exploitation of the specificities of the surgical case at hand*. We are targeting CA for AF surgeries for a custom-tuned registration algorithm, with the secondary objective of extending the algorithm to SfVM in order to assess how generalizable the method is.

Registration is a mature research field; it is necessary to survey the work of others in order to construct the knowledge repository necessary to innovate. The literature is surveyed in chapter 2 to that effect, with a special emphasis on techniques applicable to the two targeted surgeries. The scientific novelties developed to solve the registration problem as well as the methods selected in order to test our algorithms are detailed in chapter 3. The results of the experiments that validate the new methods and a discussion on their implications are in chapter 4. Chapter 5 concludes with a brief overview of the main scientific contributions as well as suggestions for future research.

CHAPTER 2 : LITERATURE REVIEW

This chapter assembles the body of knowledge necessary to build a state-of-the-art 3D-2D registration system aimed at assisting image-guided surgical interventions.

First, section 2.1 broadly covers the federation of smaller components which are the building blocks of a registration algorithm. Many of these modules are worthy of a thesis in themselves — the literature is scrutinized for widely used solutions with enough details to identify which components work best for which problem category. Section 2.2 lists the *nature* that a registration algorithm can exploit: how different data sources can be used in registration and the implication of that choice. This is where the specific surgical application for which image-guidance is most important. Since insight often comes from gleaning at work seemingly unrelated to the current problem at hand, it also broadly covers ways to deliver registration that may not be directly applicable to image guided interventions but can hint to possible innovations across applications boundary. Section 2.3 is devoted to the different validation and comparison techniques for registration algorithms. The last section (2.4) defines the research objectives.

After reading this chapter, the reader should have a good idea of the state-of-the art in 3D-2D registration algorithms, how they are built as well as the challenges to overcome in order to provide useful novel approaches to registration for image-guided surgeries.

2.1 Registration Components and Systems

Solving a 3D-2D registration problem is finding the parameters of the *transformation* that expresses the 3D and 2D data in the same coordinate frame [16]. After registration, the 3D data is aligned with the 2D data: when superposed, the intersection of the 2D plane and 3D volume map to the same physical location. In IGIs, the registration process can be seen as recovering the *pose* of the patient: finding the position and orientation of the patient in space.

The iterative registration process is as follows. First, the 3D image is initialized at a plausible starting position in space. It is then projected (2.1.1) to a 2D image, which is in turn compared to the intraoperative image from the 2D modality using a *similarity measure* (2.1.3). The similarity measure produces a cost from the images comparison, which the *optimizer* (2.1.4) uses as a basis to select the next parameters in the *transform* space. The selection of those parameters is an informed guess made by the optimizer of a position in the transform space that is plausibly closer to the

registered position than the current pose. If the optimizer judges that the current transform (2.1.2) is the registration's answer, the algorithm stops. If not, the *registration loop* is reinitialized with the new, plausibly better position and the next iteration starts. This process is illustrated in figure 2.1, and represented mathematically by equation 2.1.

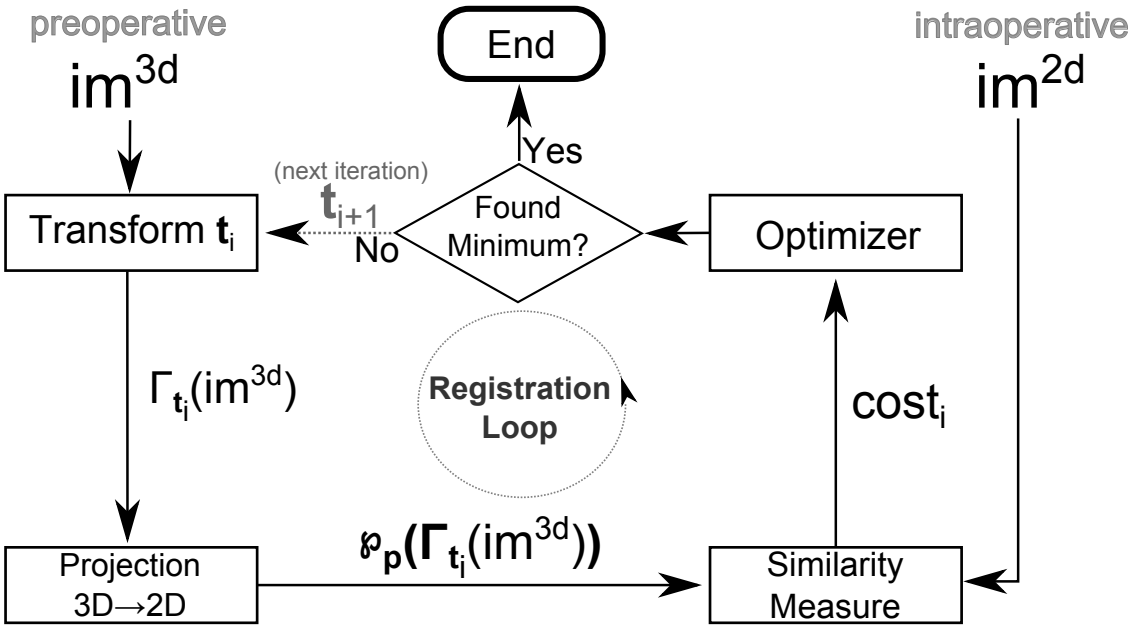


Figure 2.1 Abstract block-based representation of a 3D-2D registration system. The operator $\wp_{\mathbf{p}}(.)$ represent a projection operation (details in section 2.1.1), and the operator $\Gamma_{\mathbf{t}}(.)$ a transformation operation (details in section 2.1.2).

Figure 2.1 is a representation of how a registration system is built. It can be seen as the simple minimization of an objective function (the similarity measure):

$$\widehat{\mathbf{t}} = \arg\min_{\mathbf{t}} \ \mathrm{SM}\left(im^{2D},\ \wp_{\mathbf{p}}\left(\Gamma_{\mathbf{t}}\left(im^{3D} \right) \right) \right) \tag{2.1}$$

where the parameters $\widehat{\mathbf{t}}$ of the transform is the solution of the registration problem: the pose of the patient. The similarity measure (operator 'SM(.)') is a function that evaluates how resemblant two images are. Note that the registration process can also be a maximization ('arg max') process rather than a minimization — it changes depending on convention or the internals of the similarity measure used. In this thesis, we use the 'minimum-seeking' convention. Equation 2.1 does not explicit that the 'arg min' may or may not yield the correct answer; in most applications, the minimization process is implemented by an approximative, iterative optimization process that may get stuck in a *local minimum*. Also, the minimum of the similarity measure might not correspond

to the correct registered position. There are therefore two reasons a registration algorithm might fail:

1. The minimum cost does not correspond to the correct registered position (the similarity measure, projection or transform is not appropriate).
2. The iterative, optimizer-based minimum-seeking system does not find the minimum cost (and is unaware of its failure).

2.1.1 Projection System and Geometry

A 3D-2D projection transforms a 3D image into a 2D image. In a registration algorithm, it is used as a mean to compare data of different dimensionality. The projection can be seen as a simple function with parameters **p**:

$$im^{\wp} = \wp_{\mathbf{p}}\left(im^{3D}\right) \tag{2.2}$$

where the $\wp$ in $im^{\wp}$ denotes that the image is $\wp$rojected. This projection if composed of two operations:

1. The *geometrical mapping* of the 3D to 2D points.
2. The *functional mapping of intensities* of the set of 3D points that map to a single 2D point in order to produce a *realistic* [1] 2D image.

In a slight abuse of notation, the parameters of projection **p** in equation 2.2 are only related to the geometrical mapping, and not to the functional mapping of intensities. Those parameters exist, but since they do not change in a *family of registrations of the same case*, they are considered intrinsic to the projection function. The parameters of **p** are as follows:

$$\mathbf{p}: \ [f,\ o_x,\ o_y,\ s_x,\ s_y] \tag{2.3}$$

where f is the *focal distance*: the distance of the projection plane from the *center of projection*. $(o_x,\ o_y)$ and $(s_x,\ s_y)$ define the center and effective pixel size of the projected image respectively. The parameters of **p**, illustrated in figure 2.2, are often called the *intrinsic perspective projection parameters* [17]. The parameters are *intrinsic* in the sense that they define how the image is formed without any reference to the external world.

1. 'Realistic' here is heavily application dependent. It usually means that it resembles the 2D modality used in the operation, which may be different from the 3D modality.

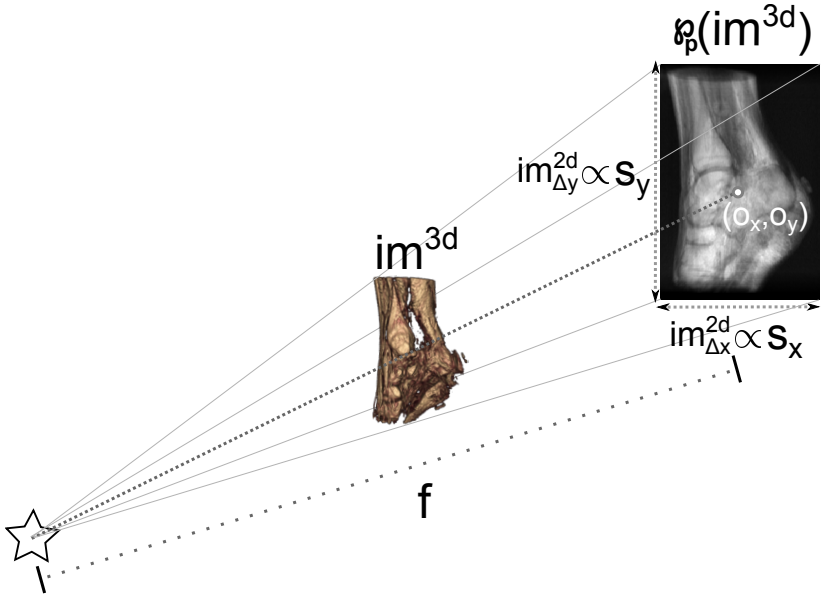


Figure 2.2 Intuitive representation of the projection parameters $\mathbf{p}$: $[f,\ o_x,\ o_y,\ s_x,\ s_y]$.

Image Formation Model

The X-ray imaging model is similar to the pinhole camera model (described in [17]), except that the image is not inverted and is magnified instead of being shrunk.

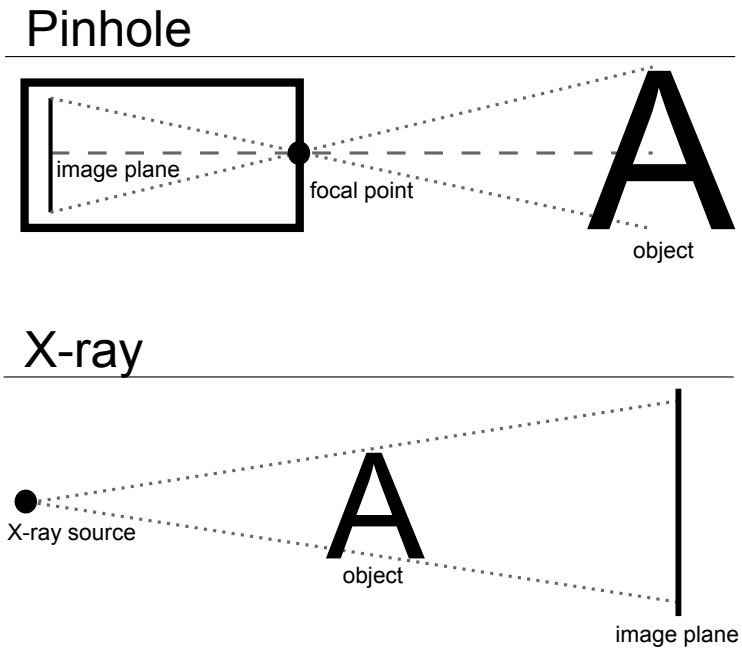


Figure 2.3 X-ray compared to pinhole projection system.

The equations modeling a perspective projection can be derived by observing the similar triangles formed by tracing rays from the *camera center* to 3D objects.

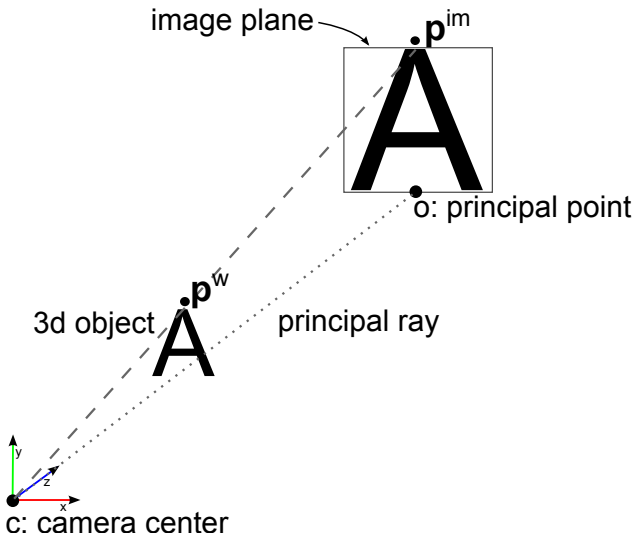


Figure 2.4 Projection system.

The points in the *world* (the 3D space) are noted $\mathbf{p}^w$ and the points on the image plane $\mathbf{p}^{im}$. Both points are 'physical units', denominated in the metric system. The *center of projection* (also sometimes called the *camera center*) '**c**' correspond to the center of projection and origin of the world coordinate system. The *principal ray* origins from the camera center and is perpendicular to the *image plane*. The *principal point* '**o**' is at the intersection between the principal ray and the image plane. The *focal distance* is measured from the camera center to the image plane and is noted 'f'. The similar triangles are best viewed when looking at the YZ plane:

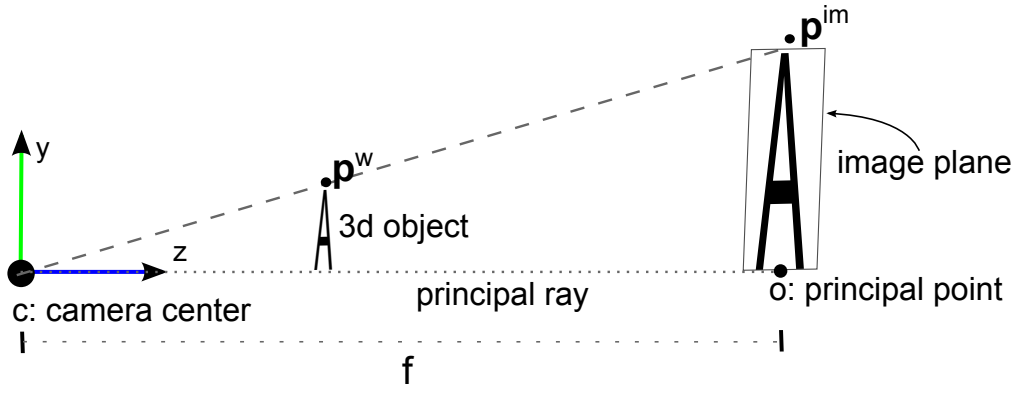


Figure 2.5 Projection system, viewed on the 'YZ' plane.

Intrinsic Parameters

The similar triangles formed by $[c_y,\ p_y^{im},\ p_z^{im}]$ and $[c_y,\ p_y^{w},\ p_z^{w}]$ allow the derivation of the formulae to project p_y^w on a 2D plane at p_y^{im}:

$$p_y^{im} = \frac{p_y^w \cdot p_z^{im}}{p_z^w} = \frac{p_y^w \cdot f}{p_z^w} \tag{2.4}$$

The same analogies are valid for the XZ plane:

$$p_x^{im} = \frac{p_x^w \cdot p_z^{im}}{p_z^w} = \frac{p_x^w \cdot f}{p_z^w} \tag{2.5}$$

Equations 2.4 and 2.5 are sufficient to project 3D world coordinates to a 2D plane. However, in order to generate a proper numerical image, it is necessary to define a scale and image boundaries. For example, if the image plane has relevant information in a zone of 1 cm by 1 cm, it could be represented either by a numerical image of 1x1 or 100x100 pixels. Additionally, the numerical image could represent just a subset of the information in the image plane.

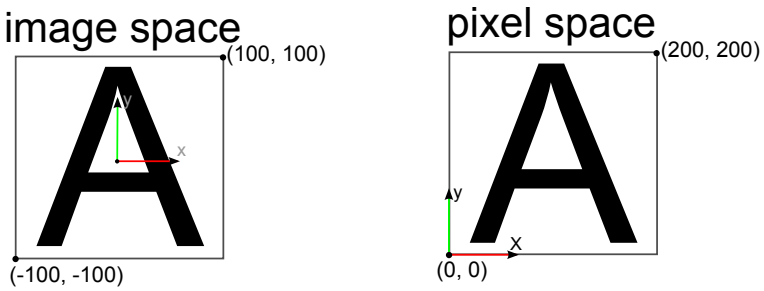


Figure 2.6 Image space to pixel space.

A scaling factor $s^{\frac{pixel}{unit}}$ defines how many pixels correspond to one *image unit* (im.u), e.g: mm. The change of referential from image space to pixel space is done by translating the scaled image units by $\mathbf{o}^{pixel}$ pixels, the position where the principal point lays in pixel space. This is done using the following equation:

$$\begin{aligned} p_x^{pixel} &= p_x^{im} \cdot s_x^{\frac{pixel}{im.u}} + o_x^{pixel} \\ p_y^{pixel} &= p_y^{im} \cdot s_y^{\frac{pixel}{im.u}} + o_y^{pixel} \end{aligned} \tag{2.6}$$

To account for image boundaries, there are two variables $\mathbf{p}^{pixel-min}$ and $\mathbf{p}^{pixel-max}$ that define the maximum and minimum points in the pixel space to be displayed in the image. Any points outside

of the minimum-maximum range created by those points are discarded. (To simplify equations, $\mathbf{o}^{pixel}$ is selected so that it lies in the middle of the image, with $\mathbf{p}^{pixel-min} = [0,\ 0]^T$.)

In order to express the projection in a matrix system, equations 2.4 and 2.5 are substituted in 2.6:

$$\begin{aligned} p_x^{pixel} &= \frac{f \cdot p_x^w \cdot s_x^{\frac{pixel}{im.u}}}{p_z^w} + o_x^{pixel} \\ p_y^{pixel} &= \frac{f \cdot p_y^w \cdot s_y^{\frac{pixel}{im.u}}}{p_z^w} + o_y^{pixel} \end{aligned} \tag{2.7}$$

which is equivalent to:

$$\begin{bmatrix} p_z^w \cdot p_x^{pixel} \\ p_z^w \cdot p_y^{pixel} \\ p_z^w \end{bmatrix} = \begin{bmatrix} f \cdot s_x^{\frac{pixel}{im.u}} & 0 & o_x^{pixel} \\ 0 & f \cdot s_y^{\frac{pixel}{im.u}} & o_y^{pixel} \\ 0 & 0 & 1 \end{bmatrix} \begin{bmatrix} p_x^w \\ p_y^w \\ p_z^w \end{bmatrix} \tag{2.8}$$

The pixel position $\mathbf{p}^{pixel}$ is recovered after division by the third component p_z^w.

This 3D-2D projection matrix is called the *intrinsic* matrix $\mathbf{M}^{int}$ because it contains the parameters that are intrinsic to the camera model: its parameters $[f,\ s_x,\ s_y,\ o_x,\ o_y]$ do not depend on variables from the *external world*.

$$\mathbf{M}^{int} \approx \begin{bmatrix} f \cdot s_x^{\frac{pixel}{im.u}} & 0 & o_x^{pixel} \\ 0 & f \cdot s_y^{\frac{pixel}{im.u}} & o_y^{pixel} \\ 0 & 0 & 1 \end{bmatrix} \tag{2.9}$$

Equation 2.9 assumes that there are no distortions in the image formation model (hence the approximation symbol). Unfortunately, X-ray systems may introduce a *shear distortion* that can be modeled by a *skew* factor (γ) [18], slightly complicating the intrinsic matrix:

$$\mathbf{M}^{int} = \begin{bmatrix} f \cdot s_x^{\frac{pixel}{im.u}} & \gamma \cdot f \cdot s_x^{\frac{pixel}{im.u}} & o_x^{pixel} \\ 0 & f \cdot s_y^{\frac{pixel}{im.u}} & o_y^{pixel} \\ 0 & 0 & 1 \end{bmatrix} \tag{2.10}$$

Extrinsic Parameters

Registration is recovering the pose of the patient, which is modeled by the *extrinsic parameters*. The *intrinsic parameters*, seen in section 2.1.1, express the internals of the *projection system*. The *extrinsic parameters* model the position and rotation of the patient in space, which is the information we wish to recover using registration. Although non-rigid deformations are plausible, this thesis is limited to the 6-dimension rigid-body parameters deformation resulting from a rotation and translation:

$$\mathbf{p}^b = \mathbf{R} \cdot \mathbf{p}^a + \mathbf{t} \tag{2.11}$$

where $\mathbf{R}$ is a 3x3 rotation matrix and $\mathbf{t}$ a 3x1 translation vector ($\mathbf{t} = [t_x,\ t_y,\ t_z]^T$). In order to store the 3D-3D rigid-body transformation in a convenient form, $\mathbf{R}$ and $\mathbf{t}$ are concatenated in a matrix with the 3D points expressed in *homogeneous* coordinates:

$$\mathbf{p}^b = \begin{bmatrix} r_{11} & r_{12} & r_{13} & t_x \\ r_{21} & r_{22} & r_{23} & t_y \\ r_{31} & r_{32} & r_{33} & t_z \end{bmatrix} \begin{bmatrix} p_x^a \\ p_y^a \\ p_z^a \\ 1 \end{bmatrix} \tag{2.12}$$

This leads to the definition of the *extrinsic matrix*, which applies a rigid-body transforming a point in the world space (3D) to another point in the world space.

$$\mathbf{M}^{ext} = \begin{bmatrix} r_{11} & r_{12} & r_{13} & t_x \\ r_{21} & r_{22} & r_{23} & t_y \\ r_{31} & r_{32} & r_{33} & t_z \end{bmatrix} \tag{2.13}$$

Projection Matrix

The intrinsic and extrinsic matrices link any 3D point in the world space to a 2D point in the image space.

$$\mathbf{p}^{2D*} = \mathbf{M}^{int}\mathbf{M}^{ext}\mathbf{p}^{3D} \tag{2.14}$$

Note that $\mathbf{p}^{2D*} \neq \mathbf{p}^{2D}$: $\mathbf{p}^{2D}$ has to be recovered by a division by the third component, as done in equation 2.8. The intrinsic and extrinsic matrices can be concatenated into a *projection matrix*, which directly links 3D and 2D positions:

$$\mathbf{M}^{proj} = \mathbf{M}^{int}\mathbf{M}^{ext} \tag{2.15}$$

2.1.2 Transform

It is the responsibility of the registration algorithm to find the right *transform* $\mathbf{t}$ that recovers the pose of the patient, which is inferred by comparing the projection of the transformed 3D pre-operative data to the 2D intraoperative images. In this process, there are a myriad different ways to transform the 3D image. $\mathbf{t}$ could be the parameters of a spline-based local deformation function applied to the 3D volume, or simple translation parameters; the selection of an appropriate transform depends of the underlying assumptions about the registration problem. If it can be assumed that the organ to be registered is solely rigidly deformed between the acquisition of the preoperative and intraoperative images, using a rigid-body 6-parameter transformation representing a 3D translation and rotation (as in equation 2.16) is adequate. These parameters are related to what is called the *extrinsic parameters* in computer vision lexicon [17], which are introduced in section 2.1.1.

$$\mathbf{t} = \begin{bmatrix} t_x & t_y & t_z & \theta_x & \theta_y & \theta_z \end{bmatrix}^T \text{ [2]} \tag{2.16}$$

Even when the rigid-body assumption is not completely correct, a registration system is often restricted to account for relatively simple transformations since there is a very high cost of raising the dimensionality of $\mathbf{t}$. If α is the dimensionality of $\mathbf{t}$, it means that the search-cost to evaluate every possible transform has a complexity of $O(n^{\alpha})$. If it requires 15 milliseconds [3] to evaluate a transform (projecting on a 2D plane and comparing to the intraoperative data), translations are evaluated in a range of 25 mm with a millimeter precision and with a 3D-angle precision of $\pm$ 1 degree, it will take 43.3 *years* to complete an *exhaustive* registration. By comparison, a similar registration but accounting for translations only (that is, $\mathbf{t} : [t_x,\ t_y,\ t_z]^T$ has a dimensionality of 3) will complete in 3.9 *minutes*. This explains why the dimensionality of $\mathbf{t}$ is often limited: to strike a precision/speed compromise.

The transform is represented by:

$$im_{\mathbf{t}}^{3D} = \Gamma_{\mathbf{t}}\left(im^{3D}\right) \tag{2.17}$$

where $\mathbf{t}$ represents the *transformation parameters* while $\Gamma_{\mathbf{t}}$ is the *function that applies the transform with parameters* $\boldsymbol{t}$ on a 3D image.

2. The capital 'T' here signifies that the vector is transposed (in order to conform to the mathematic convention of this thesis), it should not be confused with a parameter of $\mathbf{t}$.

3. This is a low-bound approximation: according to [19], generating the projected image typically takes about 15 ms when implemented on a fast GPU — this 15 ms does not include the evaluation of the similarity measure.

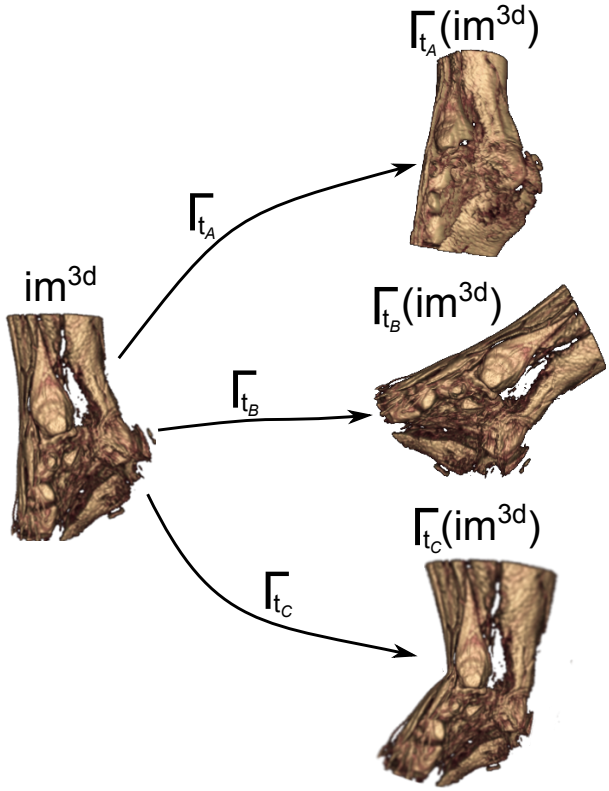


Figure 2.7 Example of transforms. $\mathbf{t}_A$ and $\mathbf{t}_B$ are rigid-body transforms, while $\mathbf{t}_C$ is an elastic transform.

2.1.3 Similarity Measure

The similarity measure 'SM', also sometimes called criterion function or cost function, measures the similarity between two images.

$$cost = \mathrm{SM}\left(im^a, im^b\right) \tag{2.18}$$

A good similarity measure has its lowest [4] cost correspond to the *registered* position of the two images, as well as smoothly and continuously decreasing as the two images are pulled from their registered positions. This is necessary since the minima is typically sought by an optimizer (see section 2.1.4) that relies on differential or local topology in order to iterate towards the minima; the presence of peaks and valleys makes it more difficult to converge towards the correct answer. The similarity measure selected is crucial for the success of the registration algorithm. Unfortunately, there exist no panacea similarity measure that is superior to all others in all aspects – it needs to be selected depending on imaging properties of the target organ, modalities used and image formation

4. Or highest depending on the convention used.

model [16]. This section briefly introduce common similarity measures and the type of data that they are expected to work well with.

Sum of Squared Differences

Perhaps the simplest similarity measure is the sum of squared difference (SSD), which compares the images pixel-by-pixel and returns the sum of the differences.

$$\mathrm{SM}^{\mathrm{SSD}}\left(im^a, im^b\right) = \frac{1}{N}\sum_{\forall(x,y)}\left(im^a_{x,y} - im^b_{x,y}\right)^2 \tag{2.19}$$

where 'N' is the number of pixels processed[5]. The SSD (equation 2.19) is intuitive, but not likely to work for multimodal image pairings: even if the images represent the same *physical reality*, the image formation model is drastically different. This means that SSD might work for 2D/2D intramodal registration, could work with quasi-intramodal registration and will most likely not be useful for multimodal or model-based registration.

Normalized Cross-Correlation

Normalized cross-correlation (NCC) measures the correlation of the deviation from the mean value of the images. This allows to register two images where the intensities are linearly correlated, thus accounting for global brightness variation.

$$\mathrm{SM}^{\mathrm{NCC}}\left(im^a, im^b\right) = \frac{1}{N}\frac{\sum_{\forall(x,y)}\left(im^a_{x,y} - \overline{im^a}\right)\cdot\left(im^b_{x,y} - \overline{im^b}\right)}{\sqrt{\sum_{\forall(x,y)}\left(im^a_{x,y} - \overline{im^a}\right)^2}\cdot\sqrt{\sum_{\forall(x,y)}\left(im^b_{x,y} - \overline{im^b}\right)^2}} \tag{2.20}$$

Local normalized cross-correlation (LNCC) is a variation of NCC that subdivides the image in blocks and computes the NCC for each of these blocks separately, then averaging the results of the individual blocks to a single scalar. This allows to account for local as well as global brightness variation in images [20].

Gradient Correlation

Gradient correlation (GC) measures the correlation between the *gradients* of two images; it detects two images as similar as long as the gradient direction and strength are linearly correlated.

$$\mathrm{SM}^{\mathrm{GC}}\left(im^a, im^b\right) = \frac{1}{2}\left[\mathrm{SM}^{\mathrm{NCC}}\left(\frac{\partial im^a}{\partial x}, \frac{\partial im^b}{\partial x}\right) + \mathrm{SM}^{\mathrm{NCC}}\left(\frac{\partial im^a}{\partial y}, \frac{\partial im^b}{\partial y}\right)\right] \tag{2.21}$$

5. The division by the total number of pixels processed is necessary since we use different image resolutions. However, it is often not required in applications where the images are always of the same size.

The derivatives in x and y are obtained by convolution with the Sobel horizontal and vertical operator respectively. GC has the advantage that it makes a less demanding claim on the resemblances of the images as NCC: as long as the two images have gradient of similar orientation and strength, they will be recognized as similar. Thus, it can be expected that GC will do better in multimodal registration than the intensity-based SSD and NCC.

Note that the gradient-based measures hinging on Sobel filters have the unfortunate property of being resolution-dependent. Great care must be taken in multiresolution schemes to adapt the kernel size to the image resolution. Also, when working with subtracted images, it might be helpful to apply 2D-preprocessing in order to remove ghost artifact whose high-gradient magnitude might adversely affect gradient-based registration [19].

Joint Entropy

Since multimodal images have a relationship that cannot be expected to be strictly linear, it is helpful to move to a more abstract level of comparison. The *Shannon entropy* $H(\mathbf{x})$ is a measure of the *quantity of information* — sometimes also seen as a measure of *uncertainty* or the *dispersion of the probability distribution* — of a signal $\mathbf{x}$. It is a weighted sum of the amount of information per outcome with the probability of that outcome.

$$H(\mathbf{x}) = -\sum_i p(x_i) \log_b p(x_i) \tag{2.22}$$

where $p(\mathbf{x})$ is the probability distribution of $\mathbf{x}$. In order to avoid infinite results, equations 2.22 implicitly defines: $p(0) \cdot log_b p(0) \to 0$. This measure of entropy can be applied to images by estimating $p(\mathbf{x})$ with the normalized histogram of the image.

$$p(im_{x,y}) := \text{normalized\_histogram}(im_{x,y}) \tag{2.23}$$

The *joint probability* of two signals $p(\mathbf{x},\ \mathbf{y})$ expresses the probability to obtain value y_i in signal $\mathbf{y}$ given x_i. It is possible to adapt equation 2.22 to compare two images to obtain the *joint entropy*:

$$H(\mathbf{x},\mathbf{y}) = -\sum_{i,j} p(x_i, y_i) \log_b p(x_i,\ y_i) \tag{2.24}$$

The joint entropy of images can be estimated by using the *normalized joint histogram* (NJH) as an approximation for $p(\mathbf{x},\ \mathbf{y})$.

$$p(im^a,\ im^b) := \text{normalized\_joint\_histogram}(im^a,\ im^b) \tag{2.25}$$

To compute the NJH, a two-dimensional array of size nBin x nBin is first initialized at 0. Then,

for all $(x,\ y)$ coordinates in $(im^a,\ im^b)$, the coordinates $(i,\ j)$ of the two-dimensional array is incremented by 1; i being the corresponding intensity of $im^a(x,\ y)$ in bin-space, and j the corresponding intensity of $im^b(x,\ y)$ also in bin-space. At the end, the resulting 2D array is normalized. Note that the bin-space is simply a linear mapping of the range of intensities in the image to $0 - nBin$, nBin being proportional to the *resolution* of the approximation of the join-probability function $p(\mathbf{x},\ \mathbf{y})$ (defines on how many integer-values the distribution is numerically represented).

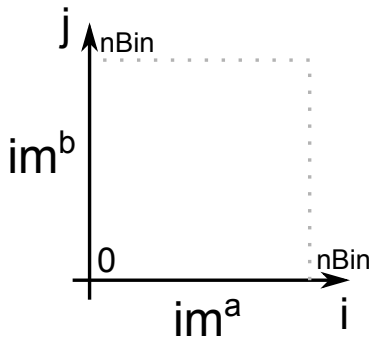


Figure 2.8 NJH(im^a, im^b) := the normalized probability that im^b has intensity j knowing that im^a has intensity i.

By substituting the joint histogram for the probability function in equation 2.24, we arrive at the final formula for the joint entropy cost function.

$$\mathrm{SM}^{JE}(im^a,\ im^b) = H(im^a,\ im^b) = -\sum_{i,j} \mathrm{NJH}(im^a,\ im^b) log_b \mathrm{NJH}(im^a,\ im^b) \tag{2.26}$$

JE is theoretically bounded from 0 to ∞ , with high values indicating that the images are not well registered. Note that since JE is a sum of $nBin^2$ normalized elements, the result is higher when using more bins in the joint histogram.

Mutual Information

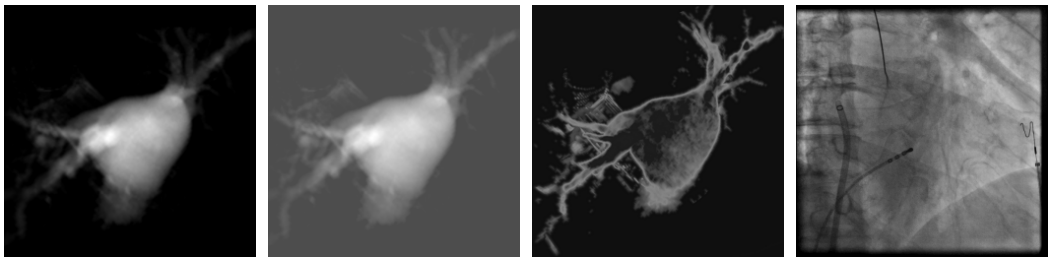
As seen in figure 2.9, the joint entropy between two dissimilar images can be low if the entropy of the different images are low as well, which could yield low results for completely misregistered images. Mutual information (MI) solves that problem by including the individual entropies in the calculation:

$$\mathrm{SM}^{\mathrm{MI}}(im^a,\ im^b) = H(im^a) + H(im^b) - H(im^a,\ im^b) \tag{2.27}$$

Equation 2.27 is bounded from $-\infty$ to 0. Another form of the metric, introduced by Studholme et al. [21], is the normalized mutual information (NMI). It has the advantage of being bounded between 0 and 1 (1: well registered) as well as being less sensitive to changes in overlap that can lead to degenerate results [6].

$$\mathrm{SM}^{\mathrm{NMI}}(im^a,\ im^b) = \frac{H(im^a) + H(im^b)}{2 \cdot H(im^a,\ im^b)} \tag{2.28}$$

6. Note that in some registration papers [21, 22], the factor two in equation 2.28 is missing, while in others it is present [23]. In our implementation, $NMI(im^a,\ im^a)$ (the maximum possible correlation: both images are the same) yields 2, which indicates that there is a factor two missing . I therefore use the version of the equation with the factor two division.

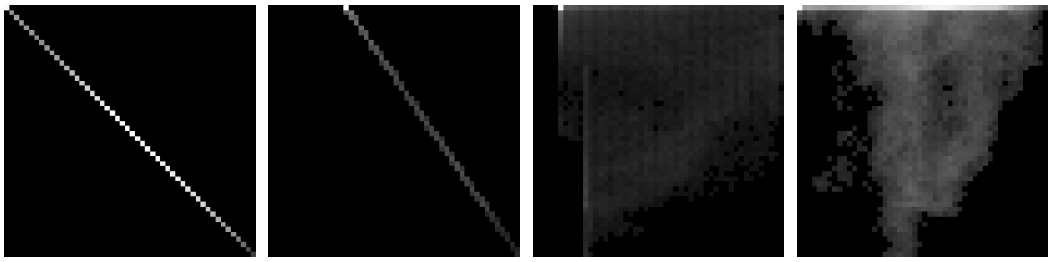


(a) DRR of the left atrium.

(b) DRR of the left atrium with higher average intensity.

(c) Maximum intensity projection of the left atrium.

(d) X-ray unrelated to previous images.



(e) JointHistogram(d,d). JointEntropy(d,d): 2.27. NMI(d,d): 1.00.

(f) JointHistogram(a,b). JointEntropy(a,b): 1.61. NMI(a,b): 0.74.

(g) JointHistogram(a,c). JointEntropy(a,c): 2.32 NMI(a,c): 0.58.

(h) JointHistogram(a,d). JointEntropy(a,d): 3.24. NMI(a,d): 0.51

Figure 2.9 The joint histograms of (d,d) and (a,b) show that when the image are linear combinations of each-other, the joint histogram has low dispersion. When the images have a multimodal correlation (a,c) or no relationship to each other (a,d), the joint histograms are more dispersed. As expected, as (a) is compared to ever-more different images, the joint entropy gets higher. Note that the joint entropy of (d,d) is higher than the joint entropy of (a,b) – this is due to (d) having a high entropy in itself, (the joint entropy of (a,a) is 1.05, as expected lower than (a,b) at 1.61). The result also show that NMI is more intuitive to use than joint entropy: exact copies of an image get a score of 1, as image get progressively less resemblant, the score monotonically lowers.

2.1.4 Optimizer

The goal of the optimizer is to find the values in the parameter space of $\mathbf{t}$ which maximizes or minimizes [7] a similarity measure:

$$\widehat{\mathbf{t}} = \arg\min_{\mathbf{t}} f(\mathbf{t}) \tag{2.29}$$

where $\widehat{\mathbf{t}}$ is the set of parameters that minimizes $f(\mathbf{t})$, and $\mathbf{t}$ are the free parameters of the trans-

7. In this thesis, the convention is optimization by minimization.

form.

Optimizers can be seen as a clever way to find the extremum of an n-dimension function *without having to evaluate all the possible positions of the parameter-space*. Most optimizers work in an iterative fashion: starting from an initial point in $\mathbf{t}$, one or more nearby positions in the parameter space are selected for evaluation. The difference between optimizer types usually lies in the selection strategy for these positions. In the set of positions (or single position) evaluated, the best approximation of the minimum value is kept as the starting point of the next iteration. Another strategy dictates when the result found is *good enough*; when it is, the optimizer returns the lowest value visited in $f(\mathbf{t})$ for all $\mathbf{t}$ evaluated. This carries the risk of returning a *local minimum*, obliviously ignoring that a lower, better value lies elsewhere.

For example, the simplest form of optimizer is an *exhaustive search optimizer*: in order to find the minimum, one can simply evaluate $f(\mathbf{t})$ for all possible $\mathbf{t}$ and return the minimum result. However, it is not practical to do so when $\mathbf{t}$ has a high dimensionality (even on modern computers, it would take *years* to find the solution to a 6-parameter registration that way).

Important Criteria of Optimizers

Optimizers can be local, semi-global or global. On one side of the spectrum, local optimizers are expected to quickly find the extremum for a simple topology but not for topologies with many creases and valleys, whereas global optimizers find the extremum even with complex topologies, at the cost of taking *much* longer to find the solution. In order to have a balance between speed, precision and robustness to complex topologies, registration algorithms predominantly use local optimizers coupled with *strategies* (see section 2.1.4) that allow them to behave in a semi-global manner.

Other important considerations for optimizers are the speed at which they can converge to the solution (how many iterations are required) as well as how long it takes to generate the candidate position(s) to be evaluated by the similarity measure. Some optimizers require the similarity measure to be differentiable analytically, whereas others do not or use numerical differentiation.

List of Commonly-Used Optimizers

Exhaustive Optimizer

It is possible to solve equation 2.29 by evaluating all possible values of $f(\mathbf{t})$ and keep the minimum.

$$\widehat{\mathbf{t}} = \min_{i=[1..n]} \left[f(\mathbf{t}_i) \right] = min \left[f(\mathbf{t}_1),\ f(\mathbf{t}_2),\ f(\mathbf{t}_3),\ ..., f(\mathbf{t}_n) \right] \tag{2.30}$$

However, this is not computationally feasible in the case of 3D-2D registration where the high-dimension of $\mathbf{t}$ makes n unfathomably large and every evaluation of $f(\mathbf{t}_i)$ is time-consuming. A more sensible solution is to sub-sample the problem space of $\mathbf{t}$ by evaluating only every m-th value in the n-dimensional search space.

$$\widehat{\mathbf{t}} \approx \min_{i=m\cdot[1..\frac{n}{m}]} \left[f(\mathbf{t}_i) \right] = min \left[f(\mathbf{t}_m),\ f(\mathbf{t}_{2m}),\ f(\mathbf{t}_{3m}),\ ..., f(\mathbf{t}_n) \right] \tag{2.31}$$

It is possible to get an approximation of $\widehat{\mathbf{t}}$ by using equation 2.31. However, there is usually a smarter way to subsample the search space by taking in consideration properties of the registration at hand. For example, Miao et al. [19] find an approximation of $\mathbf{t}$ by limiting the search to 2 dimensions in downsampled images of 64x64 pixels (the initial image have a resolution of 256x256 pixels) in a range of [-20 cm, 20 cm] x [-20 cm, 20 cm], with five discrete level of z-translation. After the exhaustive 'subsampled' optimizer finishes, a full-blown optimization is run (all parameters in $\mathbf{t}$) starting from the best value of $\mathbf{t}$ found by the exhaustive subsampled optimizer. This indicates that an exhaustive subsampled optimizer can be used to find a first approximation of $\mathbf{t}$ which can later be used as a starting point for a different optimizer type.

Gradient Descent

Gradient descent [8] consists of simply iterating towards the direction opposite to the highest gradient:

$$\mathbf{t}_{i+1} = \mathbf{t}_i - \alpha \nabla f(\mathbf{t}_i) \tag{2.32}$$

where α is the learning rate. There are many possible stop criteria, a simple one being $|f(\mathbf{t}_{i-1}) - f(\mathbf{t}_i)| < \varepsilon$ [24], where ε is a small value that is proportional to the precision required. Note that α can vary during the optimization process. The presence of the gradient implies that the similarity measure has to be differentiable.

Best-Neighbor

The best-neighbor optimizer is strikingly simple: given starting position $\mathbf{t}_1$, the neighbors at a distance of 'step size' α in the n-dimensional space are evaluated. If one of the neighbors has a

8. Gradient ascent and gradient descent are the same algorithm, excepted that the negation in equation 2.32 is switched for an addition when ascending.

lower value than the current position, the current position is moved to the lowest-cost neighbor and the next iteration starts. If the current position has the lowest value amongst all visited, the step size is multiplied by a value between 0 and 1 (the step size multiplier), and closer neighbors are explored. The stopping condition is triggered when α reaches a predefined value, for example 1.

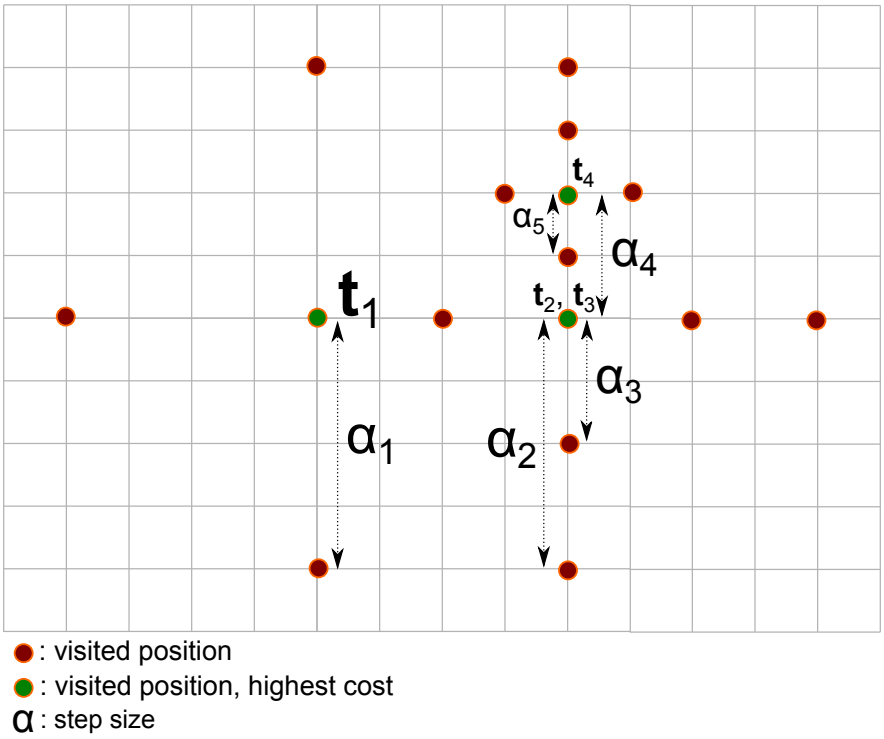


Figure 2.10 Example of a best neighbor optimizer with an initial step size of $\alpha_1 = 4$, a step size multiplier of 0.5 and minimum step size of 1. The optimizer starts at $\mathbf{t}_1$; then four positions at a distance of α_1 are visited: the lower cost is found on the right-hand side, which becomes position $\mathbf{t}_2$. From $\mathbf{t}_2$, none of the positions at a distance of α_1 have a lower cost, so $\mathbf{t}_3$ begins from the same position but with $\alpha_3 = \alpha_2 \cdot 0.5 = 2$. A higher position is found two units 'up' and the optimizer moves to $\mathbf{t}_4$. From $\mathbf{t}_4$, no higher position is found at a distance of 2, so $\alpha_5 = \alpha_4 \cdot 0.5 = 1$. The stopping condition is $\alpha = 1$, so the optimizer stops at the position $\mathbf{t}_4$.

The best-neighbor optimizer has the advantage of being simple to understand and implement. It does not require the evaluation of derivatives, therefore does not require the cost function to be differentiable. It also has few, intuitive parameters: initial step size, step size multiplier and minimum step size.

Powell

It is possible to extend a good 1-dimension optimization strategy for n-dimensions functions by optimizing one dimension at a time. However, a simple implementation will likely produce an inefficient path to the minimum. The optimization of one dimension may interfere with the optimization done for the other dimensions, which can result in an infinite loop, never minimizing the function.

Powell's method [25] is based on the idea that is it possible to find conjugate directions for which the two aforementioned problems does not occur [26]. It therefore iteratively finds a position which has a lower cost, using a set of directions which are likely to quickly lead to the global minimum.

Downhill Simplex

The downhill simplex method, also sometimes called Nelder–Mead method, downhill simplex or amoeba, generates a simplex [9] in the search space which is iteratively modified to converge towards a minimum.

The n points of a simplex in the n-dimension search space is represented by $\mathbf{P} = [\mathbf{p}_1, \mathbf{p}_2, ..., \mathbf{p}_n]$. The initial summits are selected as a rough-guess of the region in which the suspected minimum lies. After evaluating the n points, it is possible to identify the point with the associated maximum cost $\mathbf{p}_{max}$. Intuitively, the minimum is more likely to lie closer to one of the other points than the maximum. This point $\mathbf{p}_{max}$ can therefore be replaced by another point *in the surrounding* of the simplex space, with the other points indicating which region is more likely to contain the minimum. The exact heuristics to select new points are described in [27] and there is an implementation available in [26].

Comparison of Optimizer Performance

The topology of registrations' similarity measure as a function of the transform parameters is highly variable depending on the type of medical images to register. Even though some papers compare an optimizer favorably using simulated or real data, it is hard to find evidence that one optimizer works better than another when used in a new registration application.

For example, [20] measures that best-neighbor, gradient descent and Powell-Brent have similar performance when registering phantom data. Sundar et al. [28] compared best neighbor, Powell and gradient descent in registration of blood vessels in neuro-interventions: in this situation, Powell has the best performance precision and speed-wise, but they ended-up using the best-neighbor optimizer since it was more robust.

9. A simplex is the generalization of a triangle to n-dimensions.

It is therefore advisable to run tests with many optimizers and select the one that is an appropriate compromise between speed, performance and robustness for the application at hand.

Strategies Often Used to Improve the Optimizer's Behavior

The use of local optimizers, even when paired with a multiresolution approach (see section 2.1.4) in multimodal registration is not sufficient to guarantee finding the correct registered position [22]. Since the high-dimension similarity measure can form many extrema, it should be expected that the simple use of local-optimizers may not reliably find the solution. It is possible to use global optimization strategies such as simulated annealing and genetic algorithms, but they are typically not found in registration systems because of performance considerations. Jenkinson et al. report that the registration using simulated annealing takes 1 hour to complete [22], compared to 20 minutes for a 'semi-global' search. Global optimization is a bit of a misnomer since those methods make it more likely that the extremum will be found, but cannot guarantee it [22, 29].

This section contains strategies that are often used to obtain an acceptable registration result with local optimizers.

Duplicate Competing Optimizations

It is possible to lower the probability of getting stuck in a local extremum by introducing redundant optimizer passes that have slightly different starting parameters [30]. This decreases the speed of the registration, but is faster and simpler than using simulated annealing or genetic algorithms. Such a strategy is used in [19], where three sub-registrations are started from tentative positions. Once they complete, the solution for which the similarity measure is the highest amongst the sub-registrations is selected as the solution for the global registration.

In [31], three successive re-initializations of the registration result from a first pass with successively constrained translation and rotation freedom increases the percentage of registration successes.

Not Every Dimension is Equal in the Search Space

To allow for optimization of a n-dimension function, a change in any of the parameters of $\mathbf{t}$ should have a similar impact on the output of the similarity measure [20]. It is difficult to enforce these properties on a similarity measure, but impact of a change in the parameters of $\mathbf{t}$ can easily be regularized by multiplying the step size of the optimizer by a scaling factor that has a different value for each parameter in $\mathbf{t}$.

For example, translation and rotation parameters typically have different impact, so they require a different scaling value. In monoplane registration, the out-of-plane translations (parallel to the *principal ray*, as defined in figure 2.4) have a smaller impact on the projection than the two other translation parameters [28], and must therefore have a large scaling value.

The scaling values are set using common sense for the registration problem (such as in the example of translation and rotation), as well as measuring the average variation of the similarity measure's output for changes in every dimension of $\mathbf{t}$. The scaling values are modified until the variation of any of the parameters of $\mathbf{t}$ has a similar impact magnitude on the result.

Multiresolution

Because global optimizers are too computationally expensive to use, registration systems are constrained to use local optimizers, which will not succeed in finding the minimum of the similarity measure unless the topology is very simple. A multiresolution approach is a compromise between the extremes of global optimization and a 'wishful' local optimization. It is similar to the sub-sampling exhaustive optimizer seen in section 2.1.4, except that the sub-sampling is done on the image-space instead of the parameter-space.

This can be achieved by building a pyramid of sub-sampled images, finding the minimum at the lowest resolution, then starting over in a region around that minimum but at one level-down in the resolution pyramid. However, one must be aware that this makes two assumptions that are not always true [22]. First, there is no guarantee that the sub-sampling reduces the number of local maxima. Second, the global minimum at the highest resolution is not necessarily in the neighborhood of the zone where the global minimum at a lower resolution is.

However, the multiresolution approach is still widely used, indicating that even if these assumptions are not guaranteed to hold, it is an acceptable way to build a registration system that offers a good compromise between performance and precision.

Optimizer Chaining

Optimizer chaining is the use of a set of registration systems, where the output of a system is the starting point of the next, forming a chain until the last system is reached.

This can be done to use one type of optimizer that is better at iterating fast towards the solution but has a low precision as a first approximation and then another optimizer that iterates slowly towards the solution but with a greater precision as a final, 'careful' approach. Another typical use is to register first for translation parameters only, then to register for translation and rotation at the same time. The concept of optimizer chaining is widely used as it allows combining the strength

of different optimizers in a single registration system.

2.2 Registration Algorithms Categories

When building a registration algorithm, two choices have a disproportionately big impact on the outcome: the choice of registration *nature* and *strategy*. The nature of the registration is the *type* of underlying data that is used; the strategy is the *process* and *building blocks* selected. Out of the two, the nature of the registration has the biggest impact because every aspect of the system has to be built in consideration of the underlying type of data. Once the nature of the registration has been selected, it is very costly to change; it can amount to restart from scratch. This is different for strategies, which can be changed along the way with relative ease. The selection of the registration nature should therefore be done with great care.

The strategies are the different optimizers, projection and similarity measures components that can be used. The nature of the registration is more fundamental: for example, one system could use extracted points from the 3D and 2D data as a basis for a landmark-based registration while another may use the comparison of the projection of the 3D image with a 2D image through a similarity measure.

The registration nature can be separated in four broad categories: extrinsic (marker or tracking based), intrinsic (intensity, feature or gradient), model (statistical or geometrical) and hybrid (a mix of different natures in a single algorithm). This section is dedicated to surveying successful registration algorithms, sorted by nature.

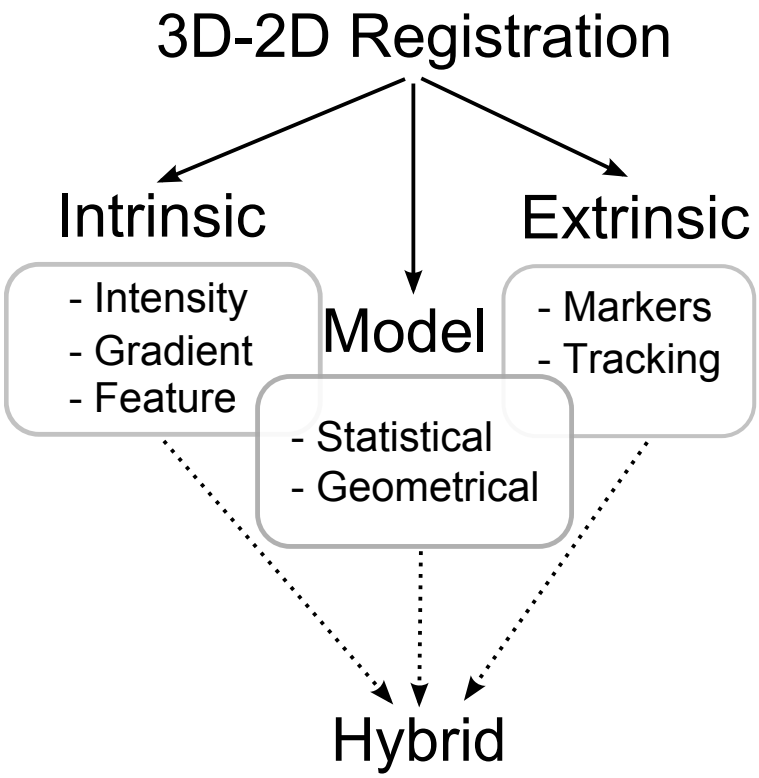


Figure 2.11 Hierarchy of different *natures* of registration. Intrinsic algorithms use information *naturally inside* the images whereas extrinsic algorithms use information artificially implanted or external to the patient. Model-based algorithms compute a model from the source data and use it as a base for registration. Hybrid algorithms use a combination of the other types of registration.

2.2.1 Intrinsic

Intrinsic registration relies solely on the information found in the medical images. It is divided in three categories: intensity, gradient and feature-based registration. *Intensity-based* registration exploits the intensities of the voxels and pixels in their integrity whereas *gradient-based* registration extracts and matches the gradients of the 3D and 2D images. *Feature-based* registration extracts features such as points or surfaces from the intensities, then use a similarity measure that solely uses the extracted features for registration.

Intrinsic Intensity-Based Methods

Intrinsic intensity-based registration (IIBR) relies solely on the voxel/pixel information to compare images. Conceptually, is it the simplest way to provide registration as it does not require the identification of special zones — images are just a *set of intensities*. The principal assumption is that it is possible to find a statistical relationship of the *degree of similarity* between the two images

via their intensities. This can be an easy task for intramodal cases where simple comparisons such as sum of square difference or normalized cross-correlation (see section 2.1.3 for an overview of these *similarity measures*) suffice, but more arduous for intermodal cases which require advanced statistical methods such as mutual information.

In order to adapt IIBR from 2D-2D to 3D-2D applications, the 3D volume is *projected* onto a 2D plane. This projection — a *simulated 2D image* — is then compared to the 2D modality. The creation of the simulated 2D image is done by tracing rays from a projection center, through a 3D volume, unto an image plane (ray-tracing). The value of each pixel can be deduced by doing an operation on all the voxels traversed by its associated ray. For example, simply summing the values followed by a re-scaling of the intensity range is sufficient to construct a 2D image that looks *something like* a projection of the 3D volume. In order to gain realism, the way the ray-tracing is done is specifically tuned to the modalities involved. For CT/X-ray, the *attenuation coefficient*, a measure of how much energy a type of tissue absorbs, as well as the logarithmic nature of the X-ray photon attenuation is modeled [32, 20]. The simulated X-ray images produced from a CT volume are called **digitally reconstructed radiographs** (DRR). They are not a completely faithful representation of a X-ray image since the effective energies of CT/X-ray are different, which result in different attenuation coefficients [33]. The modalities also have different resolutions and idiosyncratic image artifacts. Therefore, one should expect that the simulated image *resembles* the modality imitated, albeit with significant discrepancies.

It is also possible to produce simulated 2D images from a MR volume, but the difference between intermodal MR/X-ray images is greater than for CT/X-ray images. This is due to the drastically different image formation process involved in the MR/X-ray case. The anatomical features and their relative position are the same, but the way they manifest themselves on the image is completely different [34]. This may be partly overcome by the use of complex similarity measures that allows matching images that have a complex non-linear relation to each-other such a mutual information, as well as visually tuning the projection process to obtain images that are more alike the modality to imitate. However, because of the large discrepancy between the projected MR images and X-rays, this methodology is generally not suitable for registration [35]. Because it is not possible to generate DRRs from MR images [36] [10], DRR-based IIRB is not directly used to register MR to fluoroscopic data. However, it is possible to register MR to fluoroscopic data using mutual information and intensity distributions along the rays. In preliminary work, Rohlfing et al. [37] uses this method to register the MRI of a skull to two orthogonal DRR images. However, the practice of registering MR/fluoroscopic images by projecting the MR on a 2D simulated image is not widespread; there are no convincing studies that validate this methodology on many patients

10. The cited paper is correct in the sense that it is not possible to generate a DRR from an MR image, but it is certainly possible to generate a *projection*. This projection will just not be a good mimic of a radiograph.

and surgical cases. Note that the terms ‘DRR’ or ‘MR-based DRRs’ are sometimes used even if the projection is meant to simulate a MR image (a more thoughtful denomination would be ‘digitally reconstructed MR image (DRMRI)’, but is not used in the literature).

A drawback of DRRs is that they are computationally costly to produce. A typical ‘straightforward’ registration evaluates the cost function around 1000 times. Khamene et al. [20] reported that the generation of a 256x256 pixel DRR alone takes 60 ms using a 256x256x216 voxel volume, or 35 ms for when the volume is downcasted to 128x128x108 voxels. Miao et al. [19] report generating a 256x256 DRR using a 256^3 voxels volume in 15 ms when computed on the GPU of a NVidia Quadro FX 360M. Hipwell et al. [32] state that a DRR-based registration takes 5 minutes to complete, whereas Van de Kraats et al. [36] report 9 minutes for a similar algorithm.

In multimodal applications, there is sometimes no meaningful statistical relationship between the different modalities. This is the case for the LA which, when viewed under fluoroscopy, is not visible unless injected with a contrast agent. Such an X-ray image cannot be registered to an MR image of the LA using IIBR, no matter how complex the similarity measure selected [16]. To remedy to that situation, both modalities are sampled during the injection of a contrast agent that renders the LA visible for a brief time period. However, even with a contrast agent, the delineation of the organ in the fluoroscopic image is not perfect in any single frame and the organ is seen *through* the propagation of the liquid, which does not completely matches the MR image (the MR image is typically synchronized with the cardiac cycle and motion-compensated, which allows to capture the organ in its entirety). Also, the X-ray images are typically taken during the surgery, which means that they contain surgical instruments which are not seen on the MR images.

In general, IIBR is more precise than feature based methods, but has a smaller capture range [38, 35]. It has the advantage of not requiring segmentation or feature detection. IIRB is often DRR-based and used for CT/X-ray registration. For example, Miao et al. [39] use mutual information (see section 2.1.3) to match a DRR of a segmented CT of the aorta. However, they resort to a *gradient-based* method to match the spine. The gradient-based approach can be used to side-step the issues of the lack of resemblances between the projected 3D and 2D images. The next section (2.2.1) covers gradient-based registration and explains why it is often used to complement IIRB-based registration.

Intrinsic Gradient-Based Methods

The IIRB methods of section 2.2.1 have limited utility in multimodal MR/X-ray applications due to the weak similarity between the projected MR and fluoroscopic images. In some applications, the *intensity gradients* prove to have a more stable intermodal relationship. The gradient-based methods consist of extracting and matching the gradients (magnitude and direction) in the

3D and 2D images. The rays, from the X-ray source to the predominant edges on the X-ray images, are expected to be tangent to the surface of main anatomical structures in the 3D volume. The registration can be done by finding the highest gradients in the 3D image, then projecting and matching with the 2D image gradients [40] or by finding the highest gradients on the 2D image, then back-projecting through the 3D volume [35].

Markelj et al. [35] demonstrated that it is also possible to transform a 3D-2D registration into a 3D-3D problem by reconstructing a coarse 3D simulated image from multiple 2D images, then using a similarity measure based on 3D gradients matching. However, this comes with the drawback of requiring multiple X-ray images from different vintage points and well as slower registration speed.

The gradient-based methods are precise, but have low range compared to intensity-based methods [35]. They are faster than pure intensity based registration methods because only a subset of gradients needs to be considered (ray-trace only from locations on the 3D or 2D images which have the highest gradients) and most of the time-consuming volume processing can be done as a pre-operative step [35, 40]. This result in registrations completing in about 25 seconds [36]. They are especially useful in applications where the intensity similarity across modalities is weak (such as MR/X-ray fluoroscopy) and for images that have strong, well defined organ borders such as bones.

Intrinsic Feature-Based Methods

Since the performance bottleneck for most registrations is the optimization/similarity measure loop, intrinsic feature based methods can lighten the load on the optimizer by simplifying the high-complexity data (2D, 3D images) to simpler representations. The registration is then solved by the minimization of distance between detected *features* in 3D and 2D. When using the most basic feature – points – the first step is to detect them in both 3D and 2D modalities. After the detection, the 3D and 2D points are matched using close form solutions [41] or with iterative processes like iterative-closest-point [42].

For example, Sundar et al. [28] use the centerline of blood vessels in neurointerventional applications as the central feature. The vessels are detected in 3D and then reduced as a set of points. The vessels are also detected on the 2D images, but are transformed into a *distance map* to allow for a quick-lookup of the distance between the 3D projected points and the centerline of the 2D image. Ruijters et al. [43] uses a similar method to register coronary arteries (feature-based registration is popular for vessel-based images since they are generally straightforward to segment in 3D and 2D). The 3D data is segmented and projected onto a 2D distance map for every transformed visited in the optimization process. A *vesselness* filter which expresses the likelihood of a pixel to belong to a vessel-like structure is applied to the 2D images. Both features in this algorithm are non-traditional

(distance map and vesselness), but illustrates the general approach of *transforming* the 3D and 2D image into *features*, then registering by optimization of the features' similarities.

Intrinsic feature-based registration is typically fast and has a good capture range. However, the precision depends on the underlying feature extraction or segmentation methods used, which are in themselves non-trivial. The precision is typically lower than with intensity-based methods. Feature-based registration can be used to get a rough initial registration which is later refined by a more precise method.

2.2.2 Extrinsic

Extrinsic registration relies on artificial information placed either on the patient or on surgical equipment that allows recovering the pose. It can be broadly divided in two categories: *marker-based*, where markers are implanted on a patient as easily-detectable features to assist registration and *tracking-based*, which relies on the detection by an external device of the patient and equipment in the operating room in order to estimate the patient's pose.

Extrinsic Markers-Based Methods

Extrinsic registration relies on the position of markers attached on, or implanted in patients. It can be seen as special case of 'intrinsic feature-based' registration, where the features are human-made, implanted and designed to be easy to identify in both modalities. Markers are typically used in operations where there are no features that otherwise allows for registration [16]. After pairing, the markers are detected in 3D and 2D, either manually or using image processing techniques. Then, the registration is deduced by finding the rigid transformation that best projects the 3D positions of the markers on the 2D position detected. This can be done using closed-form solutions that minimize the error in the least-square sense [44], or by an iterative process [45].

For example, Simard et al. [45] uses markers taped to the skin of patients to register cases of SfVM. The markers are manually identified and paired in the 3D (MR) and 2D (fluoroscopy) modalities. Starting from an initial pose recovered from the DICOM file headers, the best transformation of the set of 3D-2D points is then found using an iterative least-square solution. The precision, measured on tests using an anthropomorphic phantom, is good (less than 2 mm). However, the pose of markers on the patient's body is time consuming and impedes the normal operation procedure. Also, the markers pairs in 2D-3D have to be identified and paired manually, which takes 2 to 3 minutes on average.

Markers-based methods have the advantage to be precise, computationally and conceptually simple,

but have the drawback that the implantation of the markers is invasive and time consuming. They are ideal to define a gold standard registration result [35], which can serve as the comparison point for other, less precise registration algorithms. It is also possible for the markers to move from their initial implantation point during surgery, causing significant registration errors.

Extrinsic Tracking-Based Methods

Tracking-based methods exploit the detection of the surgical environment via sensors to keep track of the patient's position, rather than trying to find relationships in medical images. For example, work by Rhode et al. [46, 14] track optical markers (infrared emitting diodes) placed on surgical equipment tracked using optical cameras (Northern Digital Optotrak 3020). In order to calibrate the system, markers have to be identified manually, but there is no manual intervention during the operation and no need to place markers on the patient. The precision, as measured on experiments using an anthropomorphic phantom, is of 2.4 to 4.2 mm for two-dimensional target registration error and 4.6 to 5.1 mm for integrated volume representation. This level of precision is claimed to be clinically acceptable to align the chambers of the heart. This solution is attractive because it does not impose an additional burden on the medical staff *during* the operation, but the availability of the optical tracker, the cost and the space it takes in an usually crowded operating room remains a problem.

Since this method does not require an iterative registration or the use of DRRs, it can be done in real-time. However, it assumes that there is no patient movement between the MR and X-ray machines, which means that the patient has to be properly immobilized and that a limited amount of time is allowed in between the MR scan and the operation. An image or feature-based registration could be done in addition to the tracking in order to allow for patient movement compensation.

2.2.3 Model-Based

Model-based registration consists of using an *artificial* modality derived from or representing the 3D image, then registering this *model* to the 2D images. The model is typically geometrical or statistical. For example, implants are usually designed as CAD (computer-aided design) geometrical models, which can be registered to 2D images of a patient's organ. Statistical models contain denser information of the represented organ, permitting production of realistic ray-traced projective images. When an atlas is available, it can be customized into a realistic patient-specific representation of the targeted organ in order to avoid imaging the patient under a 3D modality (to save time, cost or to avoid needlessly irradiating the patient). Fast registration can be achieved by using a model that is a *simplification* of the 3D data. This simple representation might also be easier to register than the initial 3D image when the original 3D and 2D images are intermodal. Note that

model-based registration differs from intrinsic feature-based approaches where the features are *extracted* from the 3D image and then directly matched to the same type of feature on the 2D image, such as matching point or surface sets. The models are not limited to being matched to a similar feature in 2D — they can be used along with the intensity-based similarity measures.

There are numerable cases of model-based registration applications in the literature, most of them for orthopedic use involving the pelvis [47, 48, 49, 50], femur [51, 52, 53], hip [54], knee [55, 29] or spine [56]. There is also some work involving vessels: coronary angiograms [31] and liver vasculature [57]. The registration is done using edge matching or contour-based methods [47, 53, 55, 56], iterative closest-point [51], DRR with a gradient difference metric [50], DRR with a MI-like metric [48, 49], DRR with both gradient and MI [54], binary masks [29, 31] or simple sampling of model-projected point intensities on the 2D image [57]. There seem to be no work related to model-based registration of dynamic soft tissue organs such as the heart chambers. This may be due to the fact that it is easy to register a mesh (a data structure that has information on the edge of the organ) to bones and vasculature which have clear boundaries, but harder for soft-tissue organs which do not.

Model-based registration can be used to accelerate DRR production which is the main speed bottleneck in 3D-2D registration (producing a DRR-like image from a mesh or a statistical model can be much faster than from a 3D image) or even completely bypass the production of DRR by using 'binary masks' methods. This is the approach used by Mahfouz et al. [29], where they observe the kinematic movement (under fluoroscopy) of patients that had knee replacement surgery in order to identify potential sources of joint stress. The CAD model of the implant is projected to produce a 2D binary mask as well as a 2D edge mask. A simple similarity measure that counts the superposition of the bright pixels with the masks is used. A weighted sum of the full mask compared with the inverted X-ray image and the edge mask is compared with the edge-enhanced X-ray. This extremely simple similarity measure can be used since the metallic implants appear much lighter (after inversion) than anything else under X-ray. The optimization is done using simulated annealing, which is time-efficient because the production of the masks is very fast.

Turgeon et al. [31] also bypass the production of DRRs by registering a 3D mesh of a coronary tree (simulated dataset) derived from preoperative 3D images (using the marching cubes algorithm) to a mask derived from a 2D angiogram of the heart with the limitation that the preoperative and intraoperative images must be acquired at the same phase of the cardiac cycle (end of diastole). Entropy correlation coefficient (a normalized form of mutual information) is used as the similarity measure, along with a transform (6 degrees of freedom) driven by a downhill simplex optimizer.

2.2.4 Hybrid

Hybrid registration refers to registrations that *combine* the results from different methods to take advantage of their specific strength. For example, [19] registers an aorta using a similarity measure that is the weighted sum of the similarity of the aorta (predominant in the image) as well as the coronary ostia (a much smaller feature). One method is used to steer the registration globally, while the other is used for precise alignment in the final steps of the registration. Vermandel et al. [58] register digital subtracted angiography and magnetic resonance angiography by combining intensity and feature based approaches, using the feature-based approach to speed-up the registration and the intensity-based approach to refine the solution. A similar approach is used by [40] where they use points on bone contours with an iterative-closest-point method to find a coarse registration, then switch to a gradient-based method to refine the solution. Hybrid approaches are especially common with gradient-based algorithm since they lack the range of other algorithms but are useful to refine a solution which is already close to the truth value. Registration algorithms used clinically tend to employ hybrid methods because it allows combining the strengths of different algorithms to achieve lower probability of complete failure which would require many algorithms based on different image properties to fail simultaneously.

2.3 Validation of the Registration Methods

Registration is difficult to evaluate because there is no standard dataset that can be used as a basis to evaluate all possible algorithms. There exist free-of-charge, fully calibrated multimodal datasets with ground truths [36, 16] to test and compare registration algorithms. If possible, this is the best way to evaluate registration. However, registration is applied on a wide range of medical images and modalities; those datasets are only useful to compare algorithms that are designed for applications that use medical images similar to their dataset. A lot of research effort is spent solving specific surgeries by using surgery-specific information that cannot be generalized to datasets from a different type of surgery.

In the absence of a public dataset with known ground-truths (which is most of the time the case when developing a new registration algorithm), it is still possible to validate and measure the quality of registration algorithms. This section reviews the methods that are routinely used to evaluate registration algorithms.

2.3.1 Validation of Similarity Measures

A full-fledged registration algorithm is a federation of smaller components that, in themselves, have enough complexity to motivate the creation of their own test protocols. If one of the components has defects, the whole system has little chance to work properly. This section outlines a

method to isolate the similarity measure from the rest of the registration algorithm.

A good similarity measure has the following properties [20]:

1. The topology of the evaluation of the parameter space of $\mathbf{t}$ forms an minimum [11] $\mathbf{t}_{min}$ where the registration is correct.
2. When moving away from $\mathbf{t}_{min}$ in any dimension of $\mathbf{t}$, the cost becomes higher *as smoothly and monotonically as possible*.
3. Properties (1) and (2) are robust to expected *radiometric differences* between the input images of the metric.

These properties can be assessed by evaluating the similarity measure over the whole parameter space of $\mathbf{t}$. However, since evaluating the similarity measure for the whole domain of $\mathbf{t}$ is computationally unfeasible in a reasonable amount of time and that graphical interpretation of high-dimensional data is difficult, is it preferable to evaluate the similarity measure by varying one or two dimensions at a time while fixing all others in the vicinity of the ground truth [22]. There is no formal guarantee that a similarity measure is smooth for *all* parameters when it is smooth for only one or two, but in practice it is often observed.

In order to avoid being overly verbose, the following nomenclature is adopted: the *cost function* refers to the variation of the *cost* of the similarity measure in *function* of a variation of the parameters of the transform.

2.3.2 Validation Methods without Known Ground Truth

Visual Observation

Visual observation by a medical professional is an important validation step. Since the evaluation methodologies are often application specific [16] and done on publicly unavailable images, the comparison of metrics can be of limited utility. The utility of the registration algorithm has to be evaluated by the end-users: the medical specialists themselves. To allow them to evaluate the registration algorithm, it is therefore important to provide images that present the registration result as a 3D-2D overlay. The drawback of visual observation is that it makes it difficult to compare the performance of different algorithms because of the qualitative nature of the visual observation.

Compare With Lowest Similarity Measure Cost Exhaustively Found

When unavailable, it is possible to approximate the ground-truth with the global minimum, which can be found by an exhaustive optimizer [29]. Of course, there is no guarantee that the

11. Or maximum, minimum is used here as a convention.

ground-truth coincides with the global minimum, which makes such an evaluation relatively weak. However, it can be useful to measure convergence properties of the algorithm as well as the impact of changing the similarity measure or optimizer type.

Consistency Study

Another way to run experiments without having a ground-truth is to crudely evaluate the registration solution manually, then to generate initial positions *around* that value. The registration algorithm is launched with those initial values and the results are then compared *to each other* instead of against a ground-truth [22]. If the registration algorithm is *consistent*, the different registrations results should be similar *to one another*.

2.3.3 Validation Methods with Known Ground Truth

Comparison of Transform Parameters

It might be tempting to directly compare $\mathbf{t}^{regist}$ and $\mathbf{t}^{truth}$ (the transform parameters found by the registration and the ground-truth respectively) as an error metric, but it is not an appropriate measure of a 3D-2D registration algorithm's performance. First, not all items in $\mathbf{t}$ have a similar impact. For example, out-of-plane translations in 2D projective images typically have a much smaller impact than the in-plane translations [28]. Also, the items in $\mathbf{t}$ are not necessarily independent. For example, the rotation and translation parameters in a 6-dimensional rigid body transform: since more than one set of rotation parameters $r_x,\ r_y,\ r_z$ can yield the same *final* rotation, it is not appropriate to use the direct comparison of parameters as a mean to evaluate registration quality (a *perfect registration* might yield a high error rate using that method) [36].

mPD and mTRE

The mean target registration error (mTRE) is the mean distance between the registered and ground truth points in 3D space and mean projection distance (mPD) is similar but *after 3D-2D projection* [36]:

$$\text{mTRE}(\mathbf{P},\ \mathbf{T}^{regist},\ \mathbf{T}^{truth}) = \frac{1}{n}\sum_{i=1}^{n} \|\mathbf{T}^{regist}\mathbf{p}_i - \mathbf{T}^{truth}\mathbf{p}_i\| \tag{2.33}$$

$$\text{mPD}(\mathbf{P},\ \mathbf{M}^{regist},\ \mathbf{M}^{truth}) = \frac{1}{n}\sum_{i=1}^{n} \|\mathbf{M}^{regist}\mathbf{p}_i - \mathbf{M}^{truth}\mathbf{p}_i\| \tag{2.34}$$

where $\mathbf{P} = \{\mathbf{p}_1,\ \ldots,\ \mathbf{p}_n\}$ are points where the error is evaluated. $\mathbf{P}$ can be fiducial points (although special care has to be taken not to introduce a bias [16]), or simply picked at random

inside a region of interest. $\mathbf{T}^{regist}$ and $\mathbf{T}^{truth}$ are the rigid body transformation matrices found by the registration algorithm and the ground truth; $\mathbf{M}^{regist}$ and $\mathbf{M}^{truth}$ the *perspective projection* matrices (the mPD is understood to be the result after division by the homogeneous coordinate).

Another closely related measure is the root mean square error (RMS) [31] (in its 2D and 3D forms). However, there is no practical reason to prefer RMS error to mPD and mTRE. mPD and mTRE are used in this thesis because they are more commonly seen in registration papers.

Statistically Significant Number of Registrations and Cases

The ultimate test for a registration framework is to have enough patients with known ground-truths to characterize the expected *success rate*, *precision* and *range* of the registration algorithm. This can be achieved by using the following protocol. For every patient, generate n random initial positions. These positions are derived from $\mathbf{t}_{truth}$ with a random deviation added to each component to simulate the initial difference between the patient's real position and the estimation that the registration algorithm starts with.

$$\mathbf{t}_{init} = \mathbf{t}_{deviation} + \mathbf{t}_{truth} \tag{2.35}$$

where $[\Delta t_x, \Delta t_y, \Delta t_z, \Delta\theta_x, \Delta\theta_y, \Delta\theta_z]^T \rightarrow \mathbf{t}_{deviation}$. The parameters of the deviation are generated from a bounded uniform distribution. The translation terms are bounded between $-\alpha..\alpha$ mm and the rotation terms between $-\beta..\beta$ degrees. For example, [31] uses a Gaussian distribution with $\mathbf{t}_{deviation}$ of 25 mm and 6° to corrupt $\mathbf{t}^{truth}$ into $\mathbf{t}^{init}$ in registration of coronary angiograms. Livyatan H. et al. [40] evaluate that the initial pose is within 10-20 mm and 5-15 degrees of the ground truth in cases of femur registration. The exact values depend on the specific surgery and can be evaluated by looking at recorded surgical cases. The number of registrations created is therefore the number of starting positions generated for each patient, multiplied by the number of cases available. Each registration yields an error measure, usually mPD or mTRE.

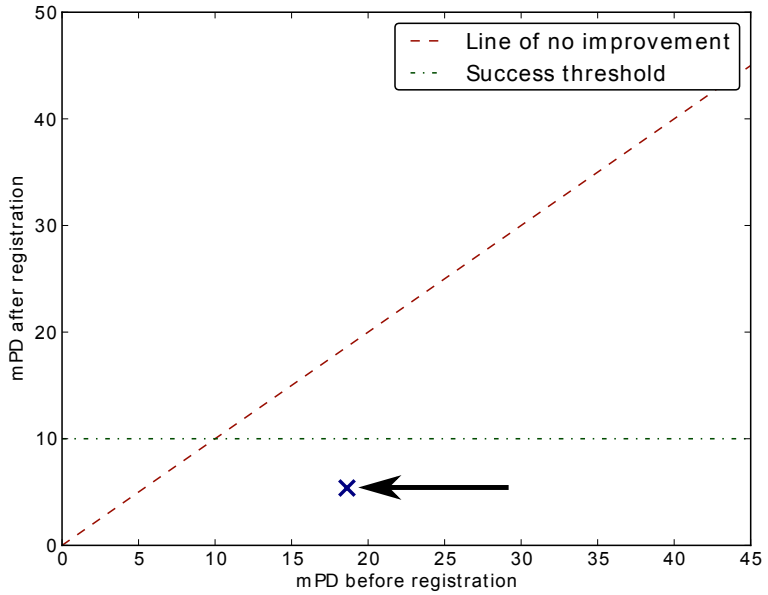


Figure 2.12 Error measurement as a function of the error before and after the registration. This example figure contains only one result, with an mPD of ≈17 mm before registration and ≈5 mm after. The success threshold is set depending on the surgery's precision requirement. The 'line of no improvement' separates the experiments where the registration improved the pose to when it was worsened.

Since a lot of data is generated from running such an experiment, it is helpful to report the result in the form of a graphic that can, at a glance, give an intuitive understanding of the properties of the registration algorithm, such as in figure 2.12 [36]. The figure expresses the relationship between the registration error before and after the registration. Every blue cross is the result of a registration. The line of no improvement separates the trials for which the registration improved the situation from those where it was worsened. The *precision* of the algorithm is the average position on the vertical axis, the *success rate* is the ratio of the cases under/over the 'success threshold' line and the *range* of the algorithm is the maximum starting error value for which *a reasonable ratio* [12] of the trials end-up under the success threshold line.

2.4 Research Direction: The Case for a New Registration Algorithm

Our goal for this research project is to provide *quick* and *automatic* registration that *does not impede the surgical procedure* for interventions where 3D-2D registration is not widely used because

12. The exact ratio to use is often a contentious debate. The surgeon is probably the best judge of what an appropriate value is.

of yet unsolved technical hurdles.

The cases of CA for AF lack predictable, distinct features that would make an intrinsic feature-based method feasible. The extrinsic methods require modifying the surgical procedure by the implantation of markers on the patient, or buying additional hardware. We saw in section 2.2 that intrinsic intensity-based methods are in general too slow for real-time registration, in large part because of the computational complexity of computing volumetric projections. This is reflected in current projection-based solutions for registration of CT/fluoroscopic data which take several minutes to complete. Also, there is currently no automatic registration algorithm that can register MR/fluoroscopic data for cases of CA for AF due to the large discrepancies between the projected MR and fluoroscopic images which even complex similarity measure based such as mutual information cannot account for. The same is true for cases of SfVM which use the MR/fluoroscopic modalities. In addition, there is a need for a common methodology that will allow registering the large diversity of medical images because VMs can be present on any body part: the organ or features to register vary from case to case.

CA of AF surgeries are currently done with an overlaid mesh that is manually registered at the beginning of the operation. We propose to use the mesh rather than the MR image to drive the registration. This is less computationally intensive since it allows side-stepping the costly volumetric projections. The mesh can be used as a way to infer zones in the fluoroscopic images, which can then be compared against each other. This also has the advantage that it abstracts the 3D modality used, since a mesh will not differ if it is segmented from MR or CT data (opening the door for fast MR/fluoroscopic registration).

We posit that the zones created by a mesh projection are most statistically different when the mesh's pose corresponds to the correct, registered position of the mesh. This is a way to solve a 3D-2D multimodal problem by the analysis of 2D-2D intramodal information, which simplifies 3D-2D multimodal analysis. At the best of the knowledge of the author, this method of registration is novel.

The **hypothesis** for this project is:

– Using the *partition* of the 2D images created by projecting the mesh derived from the 3D modality can solve a multimodal MRI/X-ray 3D-2D registration problem faster than the volumetric projection based methods currently used.

The **specific objectives** are:

I. Devise a fast, robust and automatic algorithm that provides MR/fluoroscopic registration for

cases of CA for AF without requiring additional hardware in the operating room.

a. Propose a similarity measure that directly registers a mesh to fluoroscopic images sampled during the injection of contrast agent in the left atrium.

b. Solve a multimodal registration problem by comparison of intramodal intensities groups in the fluoroscopic images.

c. Accommodate for partial match between modalities. In cases of CA for AF, the 3D image matches the contrast agent only partially (especially at the level of the PVs).

II. Build a validation framework and gather validation cases:

a. Build a tool that allows finding the ground-truth visually for the surgical cases.

b. Build a tool that allows to visually derive the biplane geometry using the source 3D and 2D images.

c. Define a preprocessing protocol that readies the patient data for registration.

d. Gather validation data from CA for AF surgeries.

III. Validate the resulting registration algorithm:

a. Validate the algorithm on surgical cases of CA for AF

b. Evaluate the algorithm on surgical cases of SfVM.

CHAPTER 3 : METHODOLOGY

There is a dearth of algorithms that allow solving hard multimodal problems (such as MR/X-ray) in an amount of time that allows for intraoperative update of the organ's pose. We introduce a new class of model-based registration algorithm which addresses both of these problems by transforming a 3D-2D multimodal problem into a 2D-2D intramodal problem. The multimodal computations are eased by first segmenting a mesh from the volumetric data, then basing the registration on the analysis of the statistical properties of the groups of pixels inferred from the projection of that mesh onto a binary mask. These operations are less costly than using volumetric ray-casting, and abstracts the type of volumetric data since the mesh not strongly coupled with the modality (e.g. CT or MR) used. These improvements come with the drawback of requiring a segmentation of the 3D volume prior to the registration. Because the technique hinges on groups of pixels derived from the projection of a mesh, we call the algorithm 'mesh-derived image partition' (MDIP) registration.

The novel MDIP algorithm is described in section 3.1. The algorithm is tested with data recorded from surgical cases which is described in section 3.2. The validation method of the proposed algorithm is presented in section 3.3.

3.1 Mesh-Derived Image Partition Based 3D-2D Registration Algorithm

The typical intensity-based approach to 3D-2D registration is to ray-cast a 3D volume to a 2D image, and then use a similarity measure to compare the projected 2D image with the 2D intraoperative frames. Based on the output of the similarity measure, the optimizer then iteratively steers the registration process. The relationship between the ray-casted and source 2D image is complex (their image formation processes are different, especially if the 3D and 2D modalities are a multimodal mix of X-ray and MR data), which requires elaborate statistical methods such as MI to compare. This results in the topology of the similarity measure as a function of a change in the transform applied to the 3D data having many peaks and creases, which makes the optimization process error-prone. This can be mitigated by using global or semi-global optimization, but at the cost of a large, negative impact on performance (discussed in section 2.1.4). The volumetric projection is also inherently slow (see section 2.2.1), which further hinders performance.

In the MDIP algorithm, a mesh is first segmented from the volumetric data, which abstracts the 3D modality used (a mesh derived from CT or MR will not have significant differences). Instead of projecting the 3D image into a simulated 2D image through ray-casting, the mesh is used to infer

groups in the 2D image, which can be done orders of magnitude faster than volumetric projection. A comparison of the *statistical properties emergent from these groups* drives the registration. Since the groups come from the same modality and only basic statistical operations are required, a lot of computation stress is removed from the similarity measure. The topology of the similarity measure's cost in function of a change in the transform's parameters is also simplified, greatly easing the optimization process.

The MDIP algorithm can be computed using the following steps. First, a mesh is used to bisect the 2D image into two *groups*: the pixels that fall under the projection of the mesh ($\in$ *mask*) and the ones that do not ($\notin$ *mask*) (see figure 3.1).

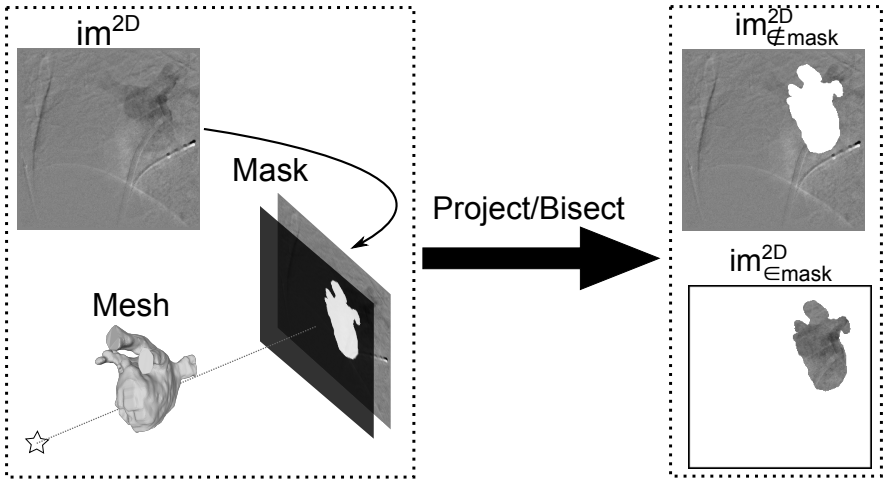


Figure 3.1 The projections of the mesh partitions a 2D image (im^{2D}) into two pixel groups: $im^{2D}_{\in mask}$ and $im^{2D}_{\notin mask}$.

The main hypothesis is that when the mesh is properly *registered*, the grouped pixels share *common characteristics* because they belong to the same entity. A similarity measure operating on *pixel groups* yields a quantity indicating how well registered the 3D and 2D images are. The set of transform parameters that correspond to the best 3D image *pose* is found by iterative optimization. The whole process is summarized in figure 3.2.

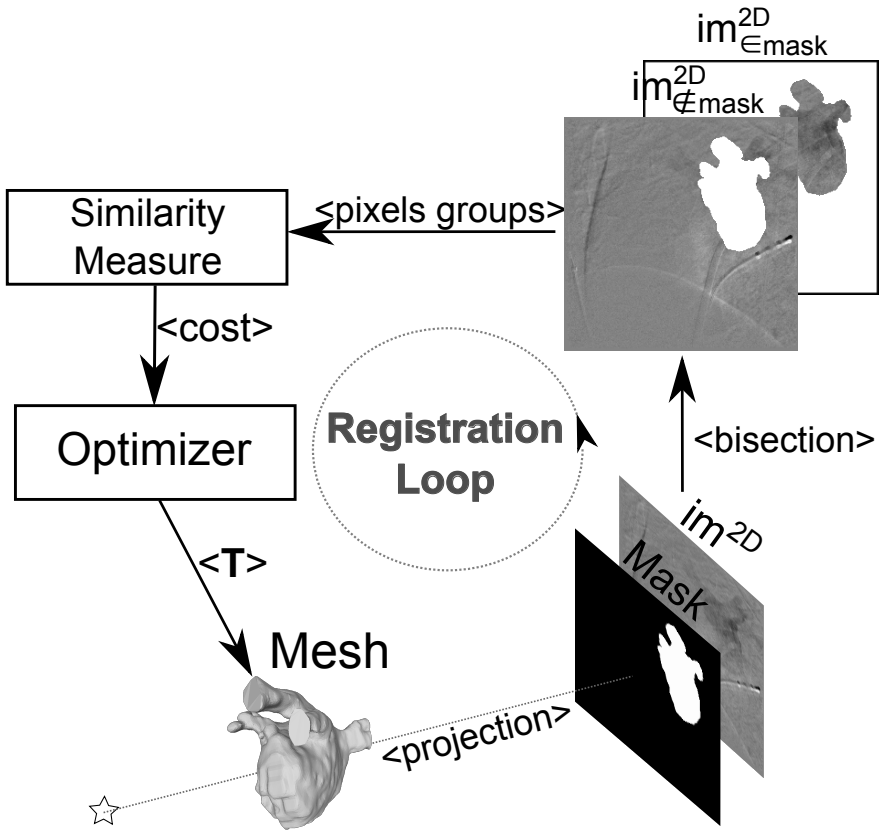


Figure 3.2 Overview of the 3D-2D MDIP-based registration algorithm.

The creation of groups in the 2D image can be seen as a way to transform the registration problem into a segmentation problem, with the possible segmentations constrained by the *topology* of the 3D image.

The rest of this section is dedicated to covering the new components that have to be developed in order to build a MDIP-based algorithm: the creation of masks from meshes (3.1.1) and the introduction of similarity measures that operate on groups of pixels with only the average intensities (3.1.2) or with histograms of intensities (3.1.3).

3.1.1 Mask Creation

In order to segment the 2D image into groups, the 3D image is projected into a mask. Although it is possible segment into groups using volumetric 3D images, the rest of the text assumes that the volumetric 3D image has been segmented into a mesh as a preprocessing step (as explained in section 3.2.3). We use meshes because they are available in the normal procedure for cases of CA for AF and they allow directly creating masks *much* faster than when using volumetric data.

The projection operation can be represented as:

$$mask := \wp_{\mathbf{p}}\left(\Gamma_{\mathbf{t}}(mesh)\right) := \text{MaskProjection}_{\mathbf{t},\mathbf{p}}(mesh) \tag{3.1}$$

where $\mathbf{t} = [t_x, t_y, t_z, \theta_x, \theta_y, \theta_z]^T$ are the extrinsic rigid-body transformation parameter and $\mathbf{p} = [f, o_x, o_y, s_x, s_y]^T$ the intrinsic perspective projection parameters (derived from the X-ray geometry parameters: $d^{source-planeA}$, $d^{source-planeB}$ and Δy^{iso}) as seen in section 3.2.2. Two OpenGL contexts (one for each plane) are created, and the meshes are loaded along with the intrinsic and extrinsic matrices which are derived from the geometry parameters previously mentioned. Figure 3.3 illustrate a simple mesh-to-mask projection system.

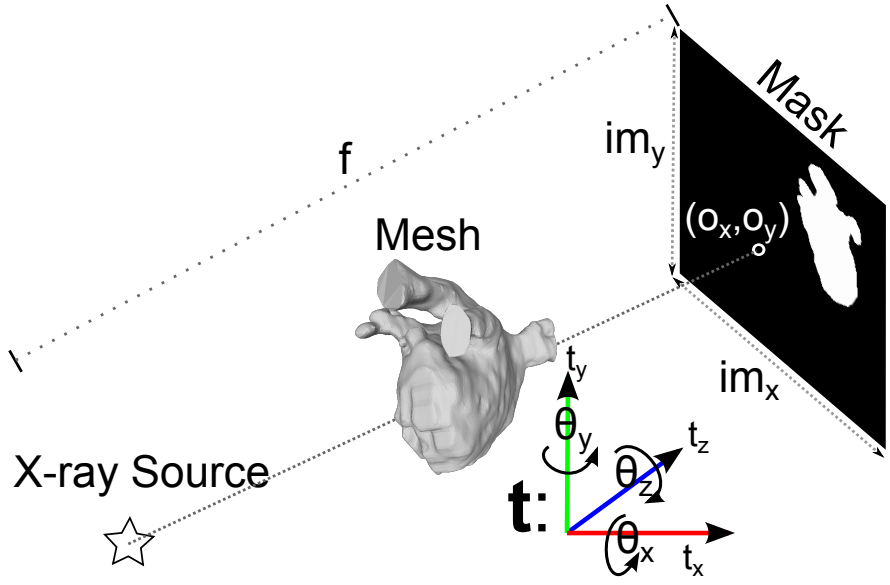
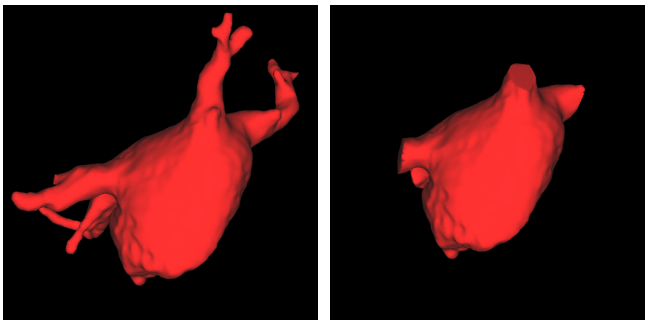


Figure 3.3 Projection system used to create a mask from a mesh.

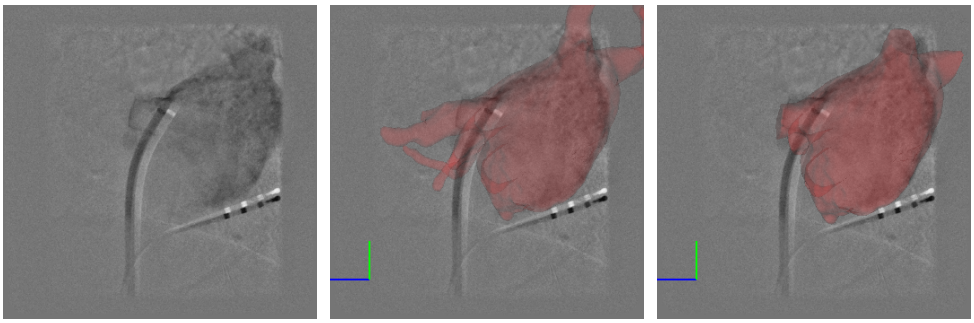
Depth-Mask Creation

Only a short part of the PVs are visible during the injection of the contrast agent, which is problematic when trying to match a mesh which has the full PVs attached.



(a) Full mesh

(b) Mesh after shortening of the PVs



(c) Subtracted fluoroscopy frame

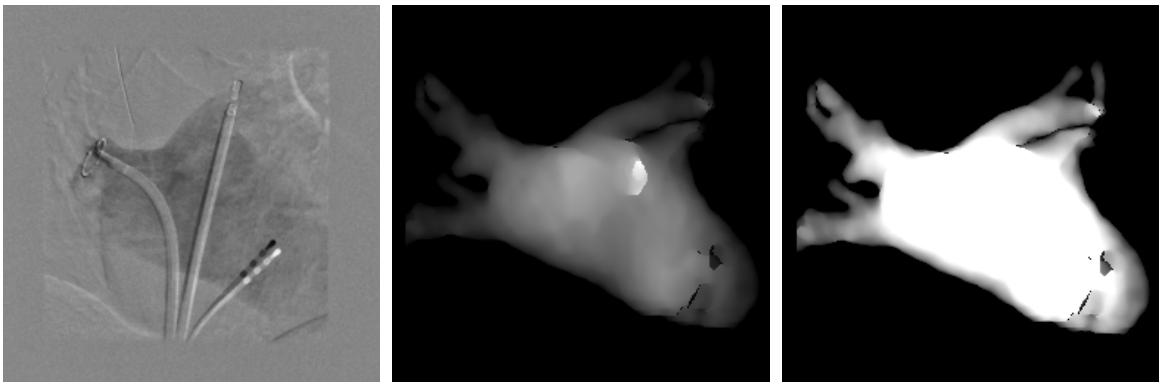
(d) Subtracted fluoroscopy frame with full mesh overlaid

(e) Subtracted fluoroscopy frame with 'shortened PVs' mesh overlaid

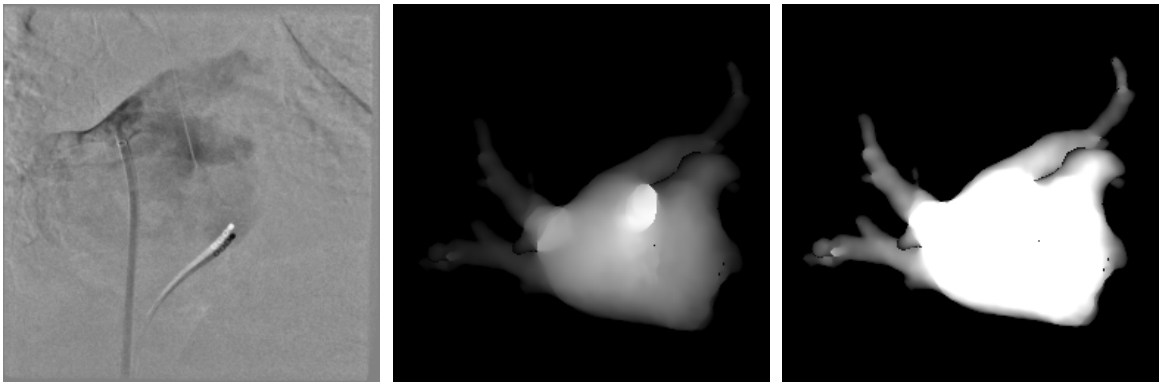
Figure 3.4 Meshes overlaid to the fluoroscopic images with PVs cut and uncut. Notice that the cut mesh matches the fluoroscopic image better than the uncut meshes.

This protruding of the PVs outside of where they are visible on the fluoroscopic images causes significant registration errors. We published a registration method that uses cut meshes such as in figure 3.4 (b) and (e) to mitigate that problem [59]. The mesh-cutting creates an additional preprocessing step which requires expertise to properly complete. The depth-mask process introduced in this section allows completely circumventing the mesh-cutting.

The main idea behind depth-mask generation is to *lessen* the importance of the pixels under the PVs since they are not visible under fluoroscopy. Since the PVs are relatively small tubes that protrude from a larger body, it is possible to produce an automatic segmentation of the mesh by using distance transform-type operations. However, because of speed concerns and simplicity of implementation, we use an *approximation of the mesh's local depth* by exploiting the depth-buffer of OpenGL. The internal depth-buffer for the front and back faces are used to approximate a ray-casting operation with close to no speed penalty over just generating binary masks (the depth-buffer is calculated by the graphics card whether or not it is used in this process) [60].



(a) P1A subtracted fluoroscopic image (b) P1A depth-mask (no processing) (c) P1A depth-mask (processed)



(d) P6A subtracted fluoroscopic image (e) P6A depth-mask (no processing) (f) P6A depth-mask (processed)

Figure 3.5 Depth-masks before and after processing. P1A refers to ‘patient #1, plane A’ and P6A to ‘patient #6, plane A’. See section 3.2 for a description of the data and associated nomenclature.

Processing is applied to the depth-masks before they are used by the similarity measure. This is done to adjust the intensity values so that pixels under the PVs have much lower values than under the LA and pixels under the LA are homogeneous. The ray-tracing approximation also produces disproportionately high values for the PVs which have their principal direction in the out-of-plane direction, as can be seen in figure 3.5 (b) and (e). This is corrected by ‘compressing’ the intensities that are more than one standard deviation over or under the mean to exactly one standard deviation over or under, then rescaling the intensities over the full range allowed by the image resolution. The result can be seen in sub-figure (c) [1] and (f). The value of 1 standard deviation (not 2, 2.5 or 3 ...) was established by a visual comparison — the value which best separates the main body of the LA and the PVs was kept.

1. Note that the depth-mask has visible artifacts for some meshes due to holes and imperfections. However, it does not seem to impede registration as the area affected is small.

Now that the *creation* process for the depth-masks is known, the next two sections describes similarity measures and how they use the information in the depth-masks (sections 3.1.2 and 3.1.3).

3.1.2 Partition-Based Similarity Measure: MeshVerage

The idea to register using pixel groups is inspired by snake methods, where a segmentation is found by iteratively evolving a curve via the minimization of an energy function. The difference in our approach is that the rigid-body parameters $\mathbf{t}$ are iteratively modified instead of the curve's control points, thus *indirectly* changing the contour of the segmentation curve according to the mesh's topology. Another way to see our solution is that it constrains the possible curves to the subset of curves that can be obtained by projecting the mesh.

If one assumes that the target 2D region is relatively homogeneous and markedly different from the other zones of the 2D image, a simple comparison of the average pixel values that fall inside and outside of the mask with the pixels in and out of these groups can be a good indication of the fitness of the registration. This is inspired by the cost function of a level-set segmentation approach introduced in [61], which leads to the definition of the following similarity measure:

$$\begin{aligned} \text{SimMeas}^{MV}(im^{2D},\, mask) := \frac{1}{n}\Bigg(& \sum_{\forall (x,y)\in mask} \left| \, im^{2D}_{(x,y)} - \overline{im^{2D}_{\in mask}} \, \right| \\ & + \sum_{\forall (x,y)\notin mask} \left| \, im^{2D}_{(x,y)} - \overline{im^{2D}_{\notin mask}} \, \right| \Bigg) \end{aligned} \tag{3.2}$$

where $im^{2D}_{(x,y)}$ is the intensity value of the fluoroscopic image at position (x,y) and $\overline{im^{2D}_{\in mask}}$, $\overline{im^{2D}_{\notin mask}}$ are the average intensity values for the *group* of pixels inside and outside the mask respectively; n is the total number of pixels processed. The pixels are scaled so that their value range from 0.0 to 1.0. The final value returned by the similarity measure is divided by the number of pixels processed, guaranteeing a result between 0.0 and 1.0. We call this similarity measure 'MeshVerage'.

Extension with Depth-Masks

In order to use the depth-masks introduced in section 3.1.1, equation 3.2 has to be adapted. The values of the depth-mask are interpreted as a *probability to belong to a group*. This probability is obtained by a simple division of the local mask value with the maximum mask-intensity value

possible (equation 3.3).

$$p_{(x,y)\in mask} = \frac{im^{2D-mask}_{x,y}}{\max(im^{2D-mask})} \tag{3.3}$$

$$p_{(x,y)\notin mask} = 1.0 - p_{(x,y)\in mask} \tag{3.4}$$

The calculation of the average values is modified in the following manner:

$$\overline{im^{2D}_{\in mask}} = \frac{1}{\sum_{\forall(x,y)} p_{(x,y)\in mask}} \sum_{\forall(x,y)} im^{2D}_{(x,y)} \cdot p_{(x,y)\in mask} \tag{3.5}$$

$\overline{im^{2D}_{\notin mask}}$ is calculated in a similar manner but with the probability $p_{(x,y)\notin mask}$ instead.

The probability is also used as a weight factor in the cost function:

$$\begin{aligned} &\text{SimMeas}^{MV-DepthMask}(im^{2D},\ mask) := \\ &\frac{1}{n} \sum_{\forall(x,y)} \left(p_{(x,y)\in mask} \cdot \left| im^{2D}_{(x,y)} - \overline{im^{2D}_{\in mask}} \right| + p_{(x,y)\notin mask} \cdot \left| im^{2D}_{(x,y)} - \overline{im^{2D}_{\notin mask}} \right| \right) \end{aligned} \tag{3.6}$$

Note that if a binary mask is used with the depth masks formulas, $p_{(x,y)\in mask}$ and $p_{(x,y)\notin mask}$ will always be equal to 0 or 1, with one being the inverse of the other, which renders equations 3.6 and 3.2 equivalent.

3.1.3 Partition-Based Similarity Measure: Histogram

The MeshVerage approach is attractive because of its simplicity, ease of implementation and speed. However, it may not be advised to discard so much of information contained in the intensities by only using averages. It is possible to extend MeshVerage by using statistical measures of information such as standard deviation and kurtosis. These values are all description of the shape of the probability density function (PDF) of the pixel intensities; rather than incrementally use more descriptive features, the *Histogram* approach uses the integral information of the discretized version of the PDF (pixel group's histogram).

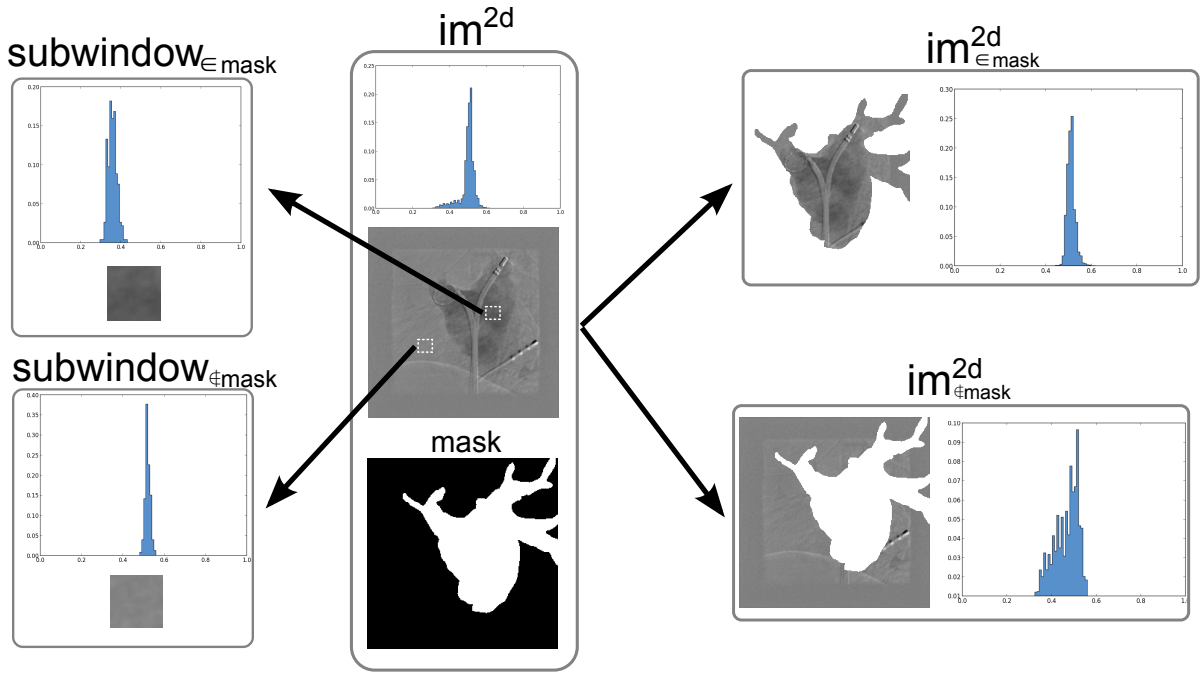


Figure 3.6 Two histograms are created: one for the pixels inside of the mask, and one for the pixels outside of the mask. Then, for every position in the 2D image, a *local PDF* is compared to either the histogram of the pixels inside or outside the mask, depending if the local position is itself under the mask or not. The *local* PDF is approximated by a local histogram formed by the pixels in a sub-window surrounding a position.

The histograms are compared by taking the absolute value of a bin-per-bin subtraction, and then averaged by the number of bins. The following equation represents how the cost is calculated using the Histogram similarity measure:

$$\text{SimMeas}^{Histogram}(im^{2D},\ mask) := \frac{1}{n}\left(\sum_{\forall(\text{subwindow})\in mask}\left|\ hist(im^{2D}_{\text{subwindow}}) - hist(im^{2D}_{\in mask})\ \right| + \sum_{\forall(\text{subwindow})\notin mask}\left|\ hist(im^{2D}_{\text{subwindow}}) - hist(im^{2D}_{\notin mask})\ \right|\right) \tag{3.7}$$

Equation 3.7 contains a division by the number of subwindows processed (n) to guarantee a result between 0.0 and 1.0.

Since it may be wasteful to compute equation 3.7 for every possible sub-window (the sub-windows of two adjacent positions have significant overlap), a *stride* parameter defines how many

pixels separate the center of two subsequent sub-windows. The similarity measure therefore has three parameters: the number of bins in the histogram, the size of the the local sub-window and the *stride* parameter.

Extension with Depth-Masks

The extension of the Histogram similarity measure for use with depth-masks is similar to what was done for the MeshVerage method (see section 3.1.2). However, since histograms are calculated instead of a scalar value (see equations 3.3 and 3.4), the calculation of $hist(im^{2D}_{\in mask})$ and $hist(im^{2D}_{\notin mask})$ are modified by including them in the sample for the histogram *a number of times* the value of $p_{(x,y)\in mask}$ or $p_{(x,y)\notin mask}$. For example, a pixel that has a $p_{(x,y)\in mask}$ of 0.9 (and therefore a value $p_{(x,y)\notin mask}$ of 0.1) is added 9 times in the array from which $hist(im^{2D}_{\in mask})$ is calculated, and 1 time in array from which $hist(im^{2D}_{\notin mask})$ is calculated.

The updated equation to take the depth-mask into account is as follows:

$$\begin{aligned}&\text{SimMeas}^{Histogram-DepthMask}(im^{2D},\ mask) := \\ &\frac{1}{n}\sum_{\forall(\text{subwindow})}\Bigg(p_{(subwindow)\in mask}\cdot\left|hist(im^{2D}_{\text{subwindow}})-hist(im^{2D}_{\in mask})\right|+ \\ &\qquad p_{(subwindow)\notin mask}\cdot\left|hist(im^{2D}_{\text{subwindow}})-hist(im^{2D}_{\notin mask})\right|\Bigg)\end{aligned} \tag{3.8}$$

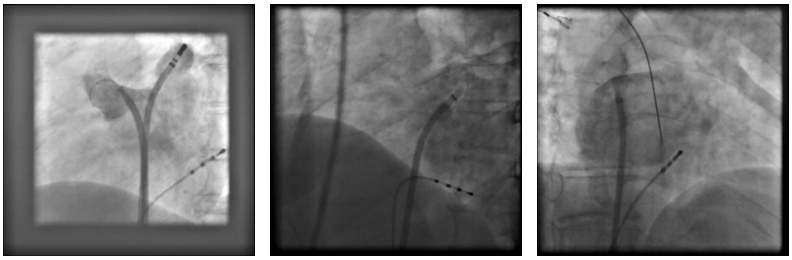
3.2 Clinical Application of Mesh-Derived Image Partition Based 3D-2D Registration

The data to register is a set of biplane cases of CA for AF from 7 patients, referred to as the 'CA for AF dataset'. The data is labeled as P#{A, B}, where '#' signifies a case number corresponding to a unique, anonymous patient. After the patient number, the associated plane, 'A' or 'B', follows. For example, 'P6B' refers to patient #6, plane B. For each patient, there is an MRI of the LA with its accompanying mesh produced by segmentation of the left-atrium from the aorta and surrounding tissue as well as a DICOM biplane fluoroscopic sequence showing the injection of contrast agent in the LA. The meshes are produced as part of the normal procedure in surgeries of CA for AF. They are overlaid to the video feed from the fluoroscope in order to help catheter guidance and display past ablation points during the operation.

3.2.1 2D Images

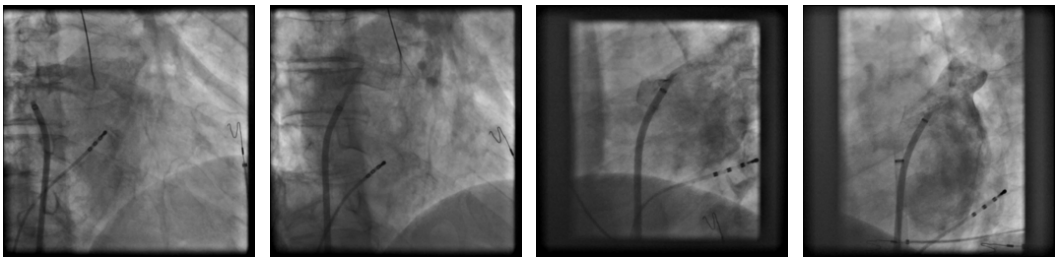
The 2D intraoperative images are sampled from a biplane X-ray fluoroscopy system. Figures 3.7 and 3.8 show selected frames in the fluoroscopic videos where the contrast agent best delineates the left atrium.

(a) P1A (b) P2A (c) P3A

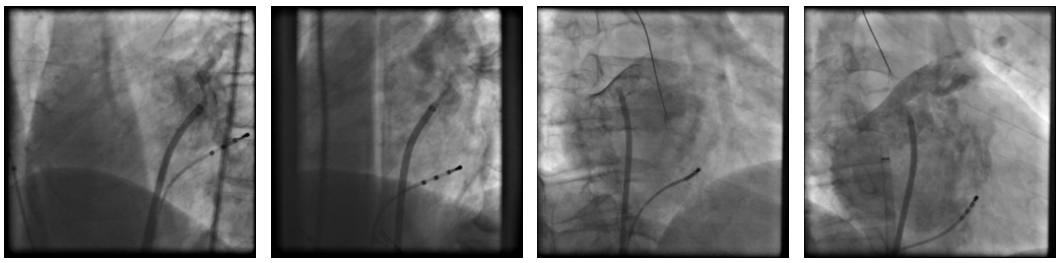


(d) P1B (e) P2B (f) P3B

Figure 3.7 Biplane fluoroscopic images for cases P1, P2 and P3 shown at the frame where the contrast agent best delineates the left atrium.



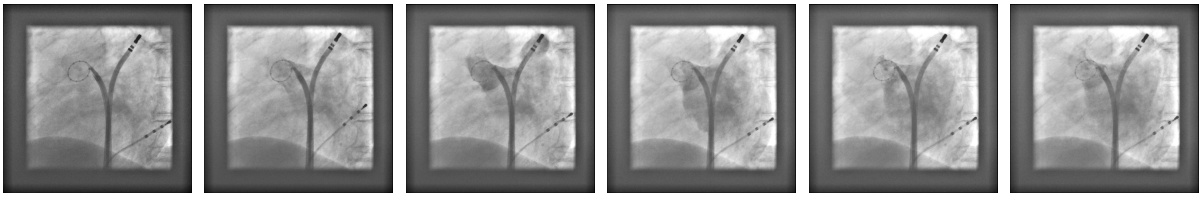
(a) P4A (b) P5A (c) P6A (d) P7A



(e) P4B (f) P5B (g) P6B (h) P7B

Figure 3.8 Biplane fluoroscopic images for cases P4, P5, P6 and P7 shown at the frame where the contrast agent best delineates the left atrium.

For each of the 7 clinical cases, there is a biplane video of between 15 to 40 gray-scale fluoroscopic frames of 1024x1024 pixels that shows the injection of contrast agent in the left-atrium. Figure 3.9 is an example of such a video. Note that the region of interest (LA and PVs) is not visible under fluoroscopy unless injected with contrast agent [3].



(a) Frame 0 (b) Frame 5 (c) Frame 10 (d) Frame 19 (e) Frame 30 (f) Frame 35

Figure 3.9 Injection of contrast agent in the left atrium for P1.

3.2.2 Geometry

In order to register biplane 3D-2D data, it is necessary to adapt the monoplane geometry introduced in section 2.1.1 for biplane cases. The surgical system's geometry is represented in figure 3.10.

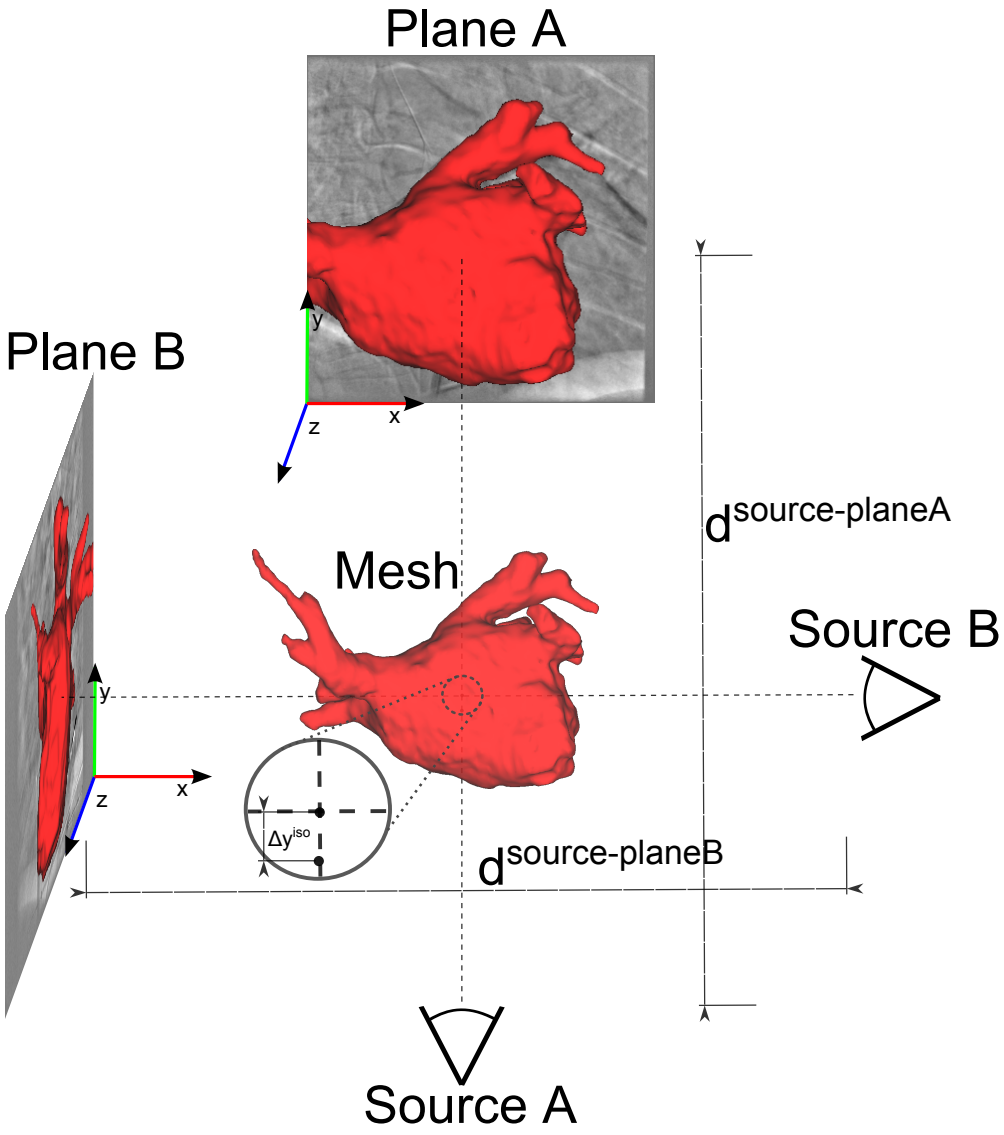


Figure 3.10 Biplane geometry for the CA for AF dataset (case P7 shown).

The angle between the principal rays of plane A and B is 90° for all cases. The parameters that vary depending on the case are:

t: The 6-dimension rigid-body transform applied on the mesh (by the registration algorithm or the user through the graphical user interface).

$\mathbf{M}^{int-planeA}$ and $\mathbf{M}^{int-planeB}$**:** The intrinsic projection parameters for plane A and B (see section 2.1.1).

$d^{source-planeA}$ and $d^{source-planeB}$**:** The physical distance (mm) between the source (X-ray source or camera center) and the imaging plane (X-ray flat panel).

Δy^{iso}**:** Physical distance (mm) on the Y axis between the principal rays for plane A and B.

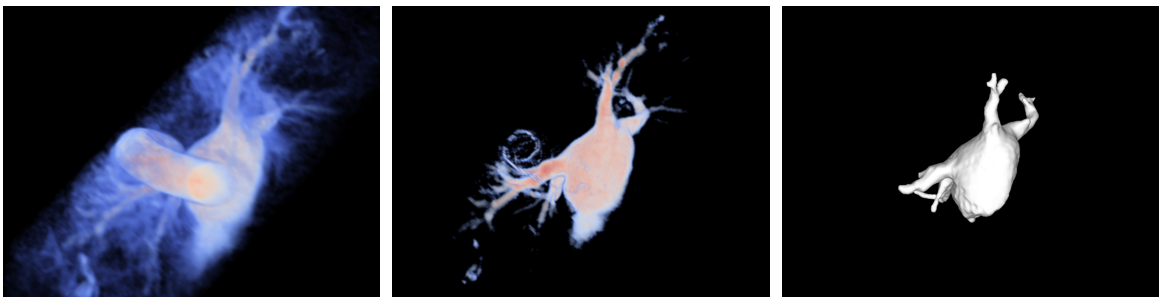
Unfortunately, the CA for AF dataset does not contain the parameters of the system's geometry. It was only available for one of the cases, which served as an initial guess for the other cases. The geometry was adjusted by visual observation with the software described in section 3.3.1. Rotations are done around the center of the mesh (calculated as the centroid of the mesh: the average position for all vertices). A number of assumptions on the constraints of the geometry are made: $d^{source-planeA}$ is forced equal to $d^{source-planeB}$ and the angle between the biplane views is fixed to 90°. These constraints are imposed because they are reasonable assumptions and manual tuning of the geometry is endless if all the parameters are allowed to change freely.

Technically, the parameters are concatenated into two extrinsic and two intrinsic matrices. Two OpenGL contexts are used to efficiently render the masks, after converting the matrices to OpenGL's idiosyncratic format [62].

3.2.3 3D Images and Segmentation into a Mesh

The experiments use 3D meshes of the LA with attached PVs. The meshes are segmented from the MR images by health-care professionals from an off-site team. Each mesh has from 15 to 20 thousand vertices, with an average of 18 thousand.

To ensure that the segmentation is reproducible (does not contain unconventional preprocessing), we compared a mesh manually extracted from the MR image using conventional procedures and off-the-shelf tools. Both meshes look similar; no significant difference was observed. Registrations done with both meshes also produce the same results. The segmentation procedure used is as follows. First, a segmentation of the MR volume is done manually. Then, the mesh is extracted by the application of an isosurface using a visually-defined threshold value. Finally, the mesh is smoothed to remove noise.



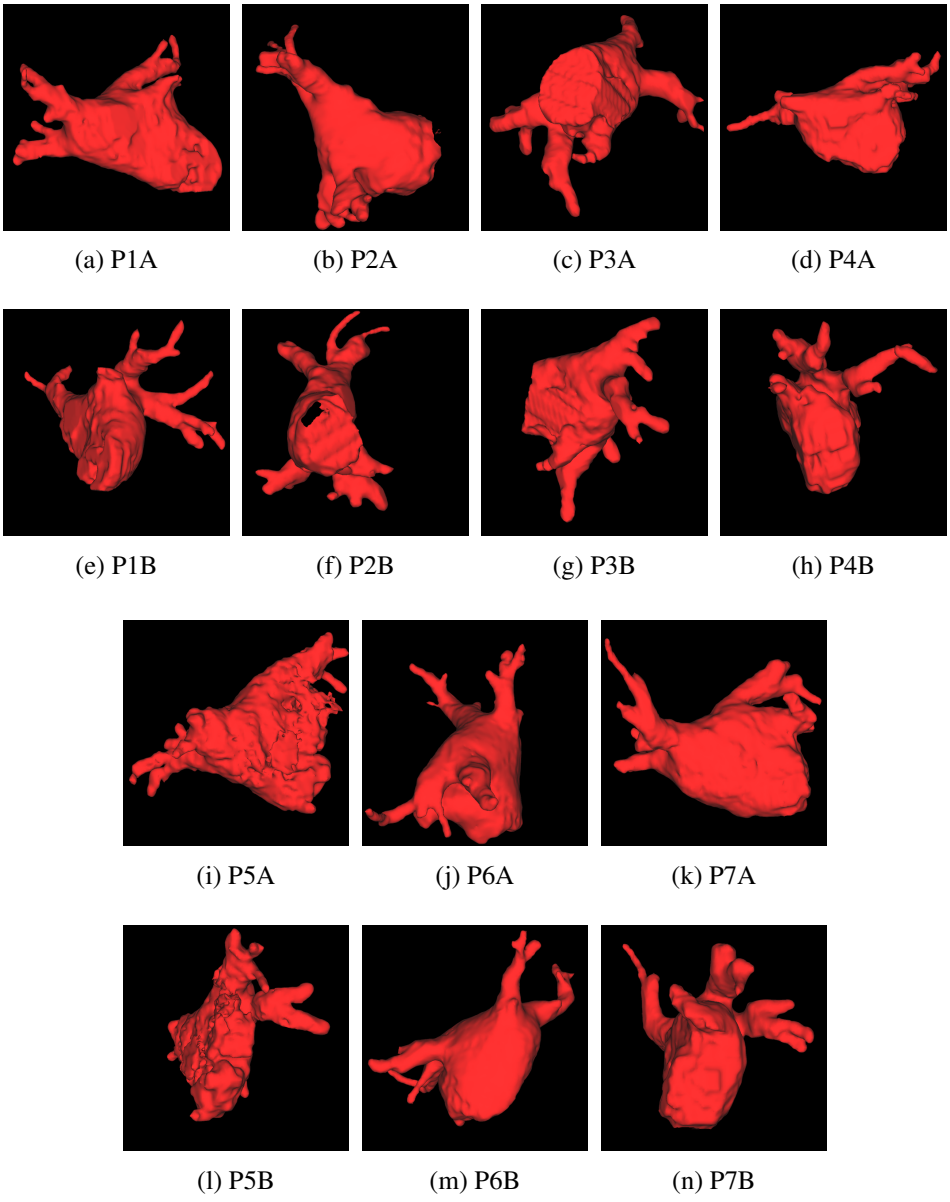
(a) P6B original MRI

(b) P6B MRI after segmentation of the LA and PVs

(c) P6B mesh used for registration

Figure 3.11 Segmentation of a mesh from a MR image.

In figure 3.11, it is worth observing that the length of the pulmonary veins in the final mesh is somewhat arbitrary. The PVs on the mesh are cut at a certain level by the segmentation, when in reality they continue into ever-diminishing capillaries. The depth-masks introduced in section 3.1.1 lessens the importance of that arbitrary choice, allowing to have the same registration result no matter the length of the PVs in the mesh. The 7 meshes used for the registration are displayed in image 3.12.



(a) P1A (b) P2A (c) P3A (d) P4A

(e) P1B (f) P2B (g) P3B (h) P4B

(i) P5A (j) P6A (k) P7A

(l) P5B (m) P6B (n) P7B

Figure 3.12 The 7 meshes after segmentation from MR images.

Figure 3.12 reveals that the mesh for case P5 contains significant noise, where the other meshes

do not. This may be due to a faulty segmentation, or maybe our collaborators simply forgot to smooth the mesh after segmentation. This has a negative impact on the registration result for this case, which is discussed in section 4.2.1.

3.2.4 Data Preprocessing

Data preprocessing in registration tend to be specific to the target surgery. Finding the right preprocessing can be guided by expertise, but there are seldom rules that dictate best-practices for all cases. A process of expertise-guided trial and error, with measurement of what helps and impedes the registration is therefore necessary, with special attention to the registration *nature* and *strategy* picked. For example, the impact of surgical instrument ghosts in digitally subtracted images might be important when using a gradient correlation similarity measure [19], but not for a similarity measure that uses average pixel values (the change in a small portion of the image does not have a significant impact on the similarity measure's output then).

This section contains experiments that indicate which preprocessing is appropriate for the task at hand.

2D Images Rescaling

The 3D image can be projected over an arbitrary-large region on 2D space. In theory, the more information in the 2D images the better. However, because it is critical to achieve fast registration and that the evaluation of the similarity measure is at minimum $O(n^2)$-bound to a 2D image of size $n \times n$ pixels, it is worth looking at the computation time / registration precision compromise.

Generally speaking, lower-resolution 2D images should result in worse registrations, but there is usually a cut-off resolution above which more information does not improve results. Since the higher resolution images come at the expense of significant processing cost, the algorithm speed is enhanced by finding the minimum resolution for which there is no precision sacrificed. In order to assess this compromise between precision and time, a monoplane registration of case P1B with the MeshVerage similarity measure (see section 3.1.2) using an exhaustive optimizer on the XZ planes (as described in section 3.3.4) is done. The registration speed is measured in *relative time to a* 1024×1024 *pixels* registration. In order to correct for registrations which end early because of erroneously terminating on a wrong solution, the time used for comparison is the average time it takes to complete a single iteration. When re-scaling is applied, it is done at the beginning of the registration using bilinear interpolation. Since the *transform* and geometry description are in physical units (mm), the resolution of the image does not necessitate a change in the registration settings.

The results and analysis of 2D image rescaling tests using this procedure are in section 4.1.1.

2D Images Subtraction

Image subtraction helps differentiating the contrast agent from the other image elements while minimizing the interference of static background elements. For example, the lower-left corner of (a), (b), (c) and (d) in figure 3.13 is much less prevalent in (e), once the images have been subtracted. However, image subtraction cause significant 'ghosts artifacts' due to the movement of the surgical instruments.

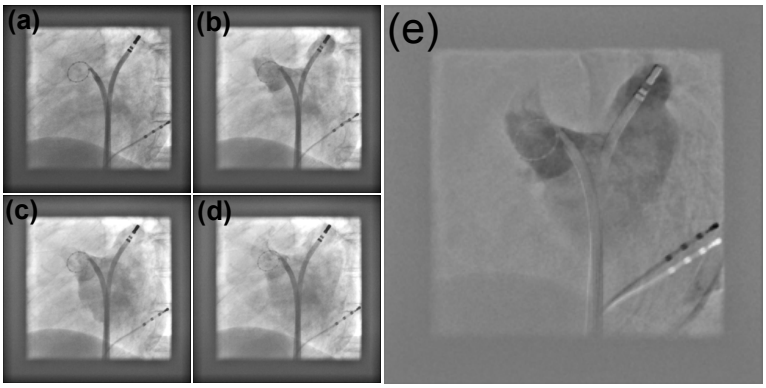


Figure 3.13 (a) to (d): Sequence of fluoroscopic images showing the injection of contrast agent in the LA (frames 0, 10, 19 and 35) for case P1. (e): Subtracted image (frame 10 - frame 0). Note that the surgical instruments are present in the images.

Registration requires the targeted organ to be visible in both modalities. Since the LA is only visible during the injection of contrast agent, the X-ray images used for registration must be sampled from the set of images containing contrast agent. This implies that the *right* frame(s) to run the registration need to be selected. The subtracted images are obtained by observation of the non-subtracted images. In the CA for AF dataset, the video sequences start just before the injection of contrast agent: frame 1 can be selected as the baseline for subtraction (frame 0 is routinely eliminated because the average intensity is often visibly different for that frame), with the other frame being the first frame that *delineates the LA well*. The selection of this second image is the only non-automated step in the registration algorithm. In order to enforce consistency, the frames subtracted are the same in the biplane sequence. For example, if frame 10 is subtracted to frame 1 for plane A, then frame 10 and 1 are also used for subtraction in plane B.

In order to evaluate if the subtraction is beneficial to the registration, the protocol established in section 3.3.4 is used with case P1B, at a 2D resolution of 256x256 pixels. The range of the trans-

form parameters t_y, t_z evaluated by the exhaustive optimizer is contained in an interval of -15 to 15 mm around the ground truth for both directions. The results are reported in section 4.1.2.

Motion Compensation. The subtracted images of section 3.2.4 have noticeable image artifacts caused by respiratory and cardiac motion between two subtracted frames. In order to determine if registrations can be improved by applying motion-compensation, the best rigid transformation possible is found by manually aligning the two images to be subtracted, as shown in figure 4.4. Then, 100 registrations with a best-neighbor optimizer, the MeshVerage similarity measure using binary masks, a random corruption (uniform distribution) of magnitude -15..15 mm for the translation in x, y and z and 10..10° for the rotation in x, y and z are computed. 50 of those registrations use the motion compensated images, 50 used the 'normal' subtracted images. The results of the analysis of the impact of motion compensation on registration using this protocol are in section 4.1.3.

3.3 Validation Methods for the Proposed Algorithm

3.3.1 Finding the Ground Truths by Visual Inspection

The ground truth is found using a custom-built program that allows modifying the parameters of the transform and geometry in a graphical user interface.

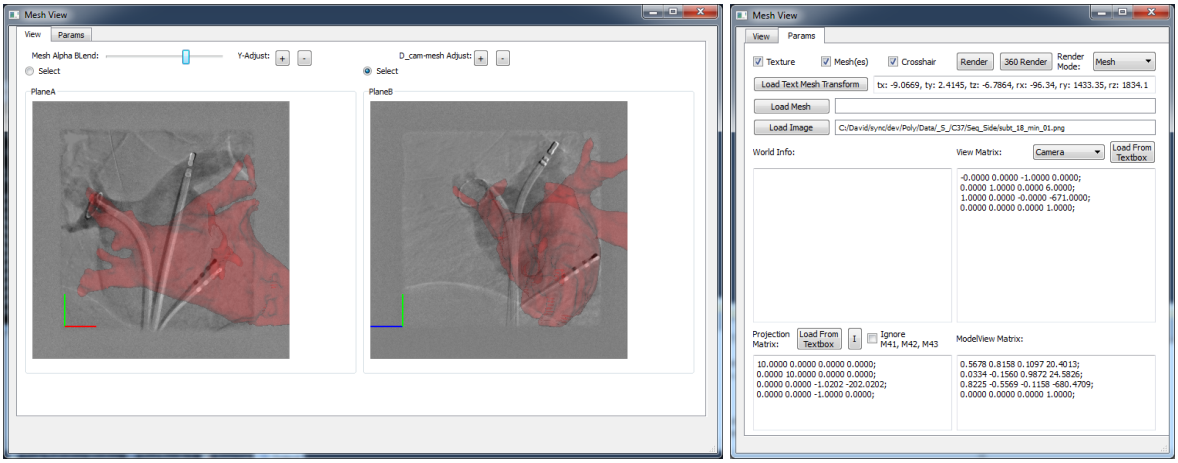


Figure 3.14 The custom-built software used in order to manually find the ground-truth values.

The software (see figure 3.14) allows defining the parameters of the biplane geometry and manipulating the translation and rotation applied to the mesh using the mouse. The left and right planes are linked: a movement of the mesh in the left plane results in a movement in the other plane according to the underlying geometry. Unfortunately, we could not get a medical specialist to evaluate the ground truth: they were found by careful examination of the MRI, X-ray and mesh

data by non-medical staff (me). In all cases, the probable positions were fastidiously evaluated and the best match kept as the ground truth. Fortunately, since all cases are biplane, it makes it less likely to get the position completely wrong: the place where the mesh is supposed to be is obvious in most cases where the contrast agent clearly delineates the LA and PVs. However, there are some cases where the contrast agent only partially render the LA and PVs visible, there remains some ambiguity on whether or not the defined ground truth correspond to the exact registered position.

3.3.2 Performance Evaluation: Precision

In order to validate the registration algorithm, the following protocol is used:

1. Find the ground truth ($\mathbf{t}_{truth}$) by visual observation using the custom-built software described in section 3.3.1.
2. Generate the initial positions ($\mathbf{t}_{init}$) by adding a random deviation vector to the ground truth. The deviation vector is $\Delta\mathbf{t}_{deviation} = [\Delta t_x, \Delta t_y, \Delta t_z, \Delta\theta_x, \Delta\theta_y, \Delta\theta_z]^T = [-15..15, -15..15, -15..15, -10..10, -10..10, -10..10]^T$ where 'A..B' signifies a random number between A and B following a uniform distribution.
3. Run the registrations using the initial positions as starting points.
4. Compare the final positions of the mesh with the ground truths. This is done both visually using the mesh contours and quantitatively using the mPD error metric.

A 6-dimension rigid-body transformation ($\mathbf{t} = [t_x,\ t_y, t_z,\ \theta_x,\ \theta_y,\ \theta_z]^T$) along with a best neighbor optimizer is used with the following parameters:

Table 3.1 Best-neighbor optimizer parameters.

Parameter	Value
Step Size for Translation	5.0
Step Size for Rotation	0.25× Step Size for Translation
Step Size Multiplier	0.5
Minimum Step Size	1.0

For the histogram-based similarity measures, the following parameters are used:

Table 3.2 Histogram similarity measure parameters. Refer to section 3.1.3 for the definition of the parameters.

Parameter	Value
Number of bins in histogram	200
Size of local sub-window	13x13 pixels
Stride	2

The first pass of the optimizer runs using translation only with the maximum step size; every subsequent pass uses all the transform parameters. The images used are subtracted but not motion-compensated. The 2D images are resized to 256x256 pixels at the beginning of the registration. The experiment generates and runs 200 registrations for every one of the 7 clinical cases. 1400 registrations are therefore run for every similarity measure, for a total of 5600 registrations.

Section 2.3.3 covered the mTRE and mPD error measures. In a previous iteration of the research work for this thesis [59], we reported the error in both mPD and mTRE. Having the two error measures did not bring additional insight (mTRE is usually a little bit higher than mPD, but follows the same pattern as mPD for our surgical cases) and made error analysis more difficult since there is always two values to consider. In addition, 3D-2D registration applications are better measured with 2D error (mPD) than 3D error (mTRE) [36] since the operation itself is conducted under the 2D modality. Because of these elements, this experiment reports error solely using the mPD measure. Instead of choosing fiducial points for evaluation, as done for other papers evaluating 3D-2D registrations for CA of AF [9], we measure the error on *all vertices of the mesh*[2]. This penalizes us since the overall error is much larger than the 'application error' (this is discussed in section 4.2.3), but allows for automatic error evaluation of the 5600 registrations. The results of this experiment are reported in section 4.2.1.

The mPD error measure is good to evaluate the whole system, but when developing the *components*, it is helpful to get insight by looking at the interim-results inside the algorithm. Section 3.3.4 describes the method that was used to get an intuitive measure of the interplay of the similarity measure, optimizer and transforms.

3.3.3 Performance Evaluation: Time

The performance of our algorithm is evaluated by automatic timers running during the registration and kept in a registration log file. After all the registrations have run, the time statistics

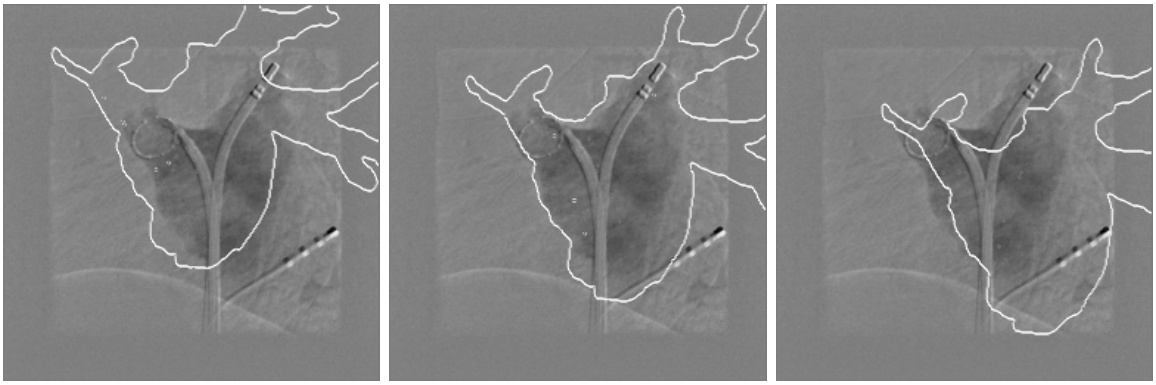
2. The average number of vertices for the left atrium meshes is around 18 thousand.

(mean $\pm$ standard deviation) are gathered and grouped by similarity measure used. The 2D image size used is 256x256 pixels, the same size used in the precision evaluation (3.3.2). The results are reported in section 4.2.2.

3.3.4 Inspection of Similarity Measures

Section 2.3.1 motivated the need for the isolation of the similarity measure from the rest of the registration system by producing a plot of the effect of changing two parameters of the transform on the similarity measure's output. There are some practicalities to consider in order to produce the cost function figures with the CA for AF dataset: a variation of the $(X,\ Y)$ axes in plane A will result in an out-of-plane translation for plane B. Since it is desirable to measure the effect of varying a single parameter at a time, the registration is run in monoplane mode.

An *exhaustive optimizer* is used to evaluate all the positions within a range and resolution around the ground truth. See figure 3.15 for an example.



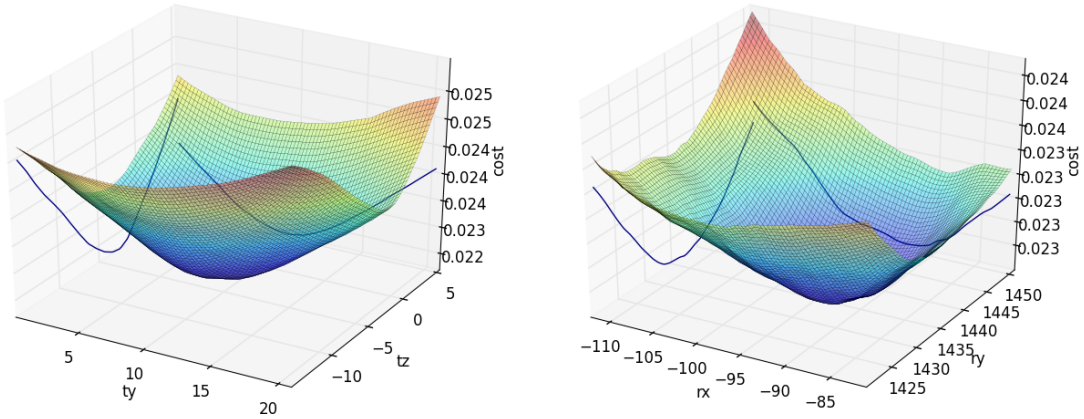
(a) P1 B at $(t_y,\ t_z)$ at $(10,\ 10)$ mm of the ground truth.

(b) P1 B ground truth.

(c) P1 B at $(t_y,\ t_z)$ at $(-10,\ -10)$ mm of the ground truth.

Figure 3.15 Exhaustive search from $(-10,\ -10)$ mm to $(10,\ 10)$ mm around the ground truth. Note that the parameters in $\mathbf{t}$ that correspond to the 'up-down' and 'left-right' translations depend on the case geometry. Refer to figure 3.10 for the cases' geometry.

Here is a graphical representation of the variation of the output of the similarity measure as a function of parameters in $\mathbf{t}$ ($\theta_x,\ \theta_y$ and $t_y,\ t_z$).



(a) Variation of $t_y,\ t_z$ in $(-10..10)$ mm.

(b) Variation of $\theta_x,\ \theta_y$ in $(-15..15)$ degrees.

Figure 3.16 Plot of similarity measure's output for case P1B un function of a variation of the parameters of $\mathbf{t}$. The curves in the background show the variation of the similarity measure's output as the transform parameters get further away from the global minima in a direction parallel to the axis.

CHAPTER 4 : RESULTS AND DISCUSSION

This chapter contains experiments that measure the performance of the MDIP-based registration algorithm proposed in this thesis. First, section 4.1 contains experiments results used to *tune the parameters* of the registration algorithm. Then, the whole registration system is evaluated on clinical cases of CA for AF in section 4.2 and SfVM in section 4.3.

4.1 Impact of Data Preprocessing

Besides the *nature* and *strategy* of the registration algorithm, there are many parameters that influence the ultimate outcome of the registration. This section contains the experiments conducted in order to determine appropriate parameters.

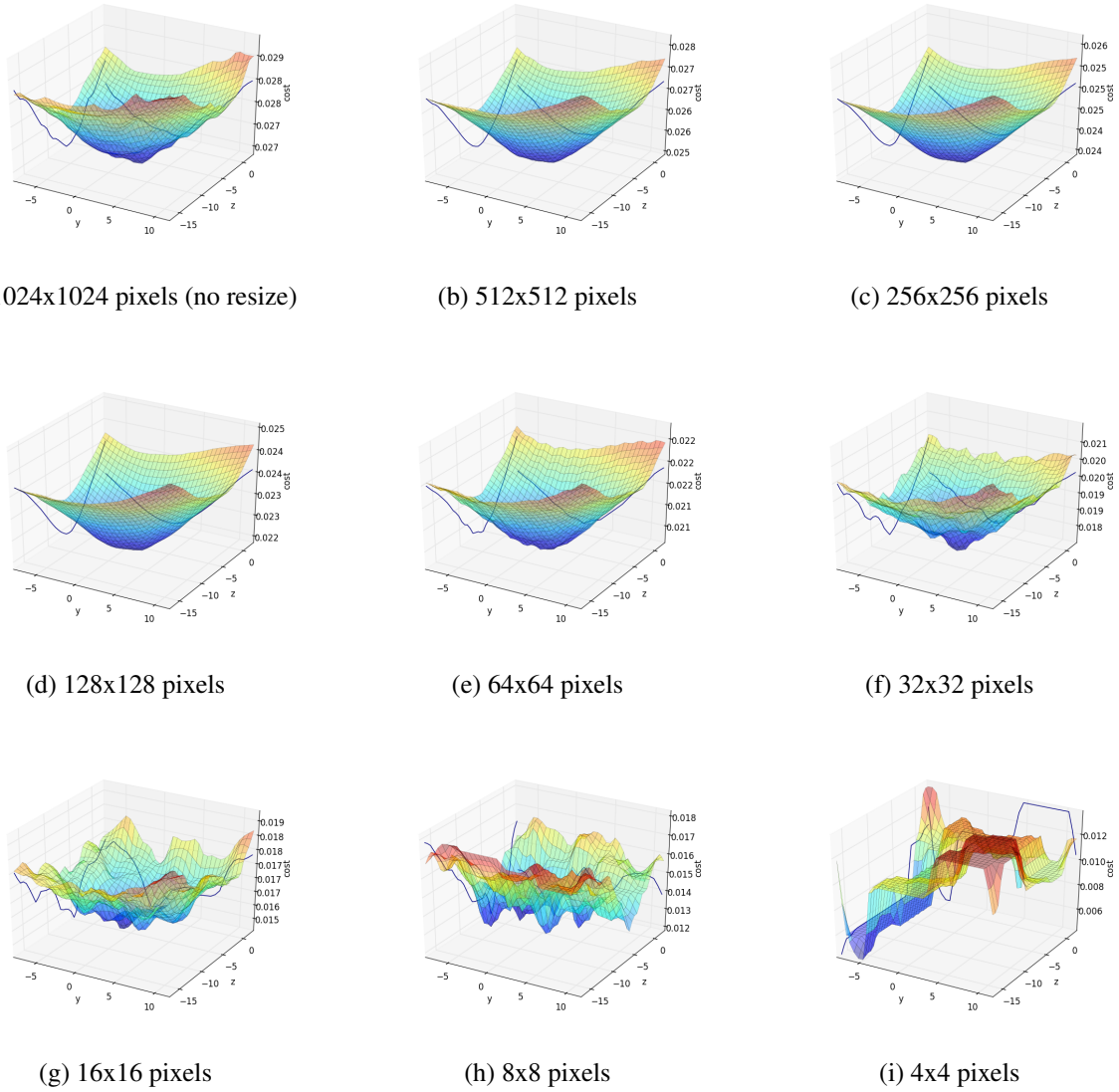
4.1.1 Impact of 2D Images Rescaling

Table 4.1 Impact of 2D image resizing prior to registration on registration precision and speed, using the MeshVerage similarity measure and case P1B in monoplane mode.

im^{2D} size (pixels)	registration speedup	mPD (mm)
1024x1024	1.00	1.34
512x512	2.15	2.67
256x256	2.72	1.89
128x128	3.10	1.89
64x64	3.24	1.89
32x32	3.37	1.34
16x16	3.34	5.97
8x8	3.11	8.96
4x4	3.20	17.11

Table 4.1 reveals that there is not much speed to be gained by reducing the image below 256x256 pixels. It seems like it is possible to register with the same precision at resolutions as low as 32x32 pixels. This may be due to the fact that the MeshVerage similarity measure is by nature based on averages, and that the number of pixels in the sample does not need to be very high in order to differentiate between the pixels inside and outside the target zone. Note that the mPD for 512x512 pixels has a higher error than for lower resolutions. However, a difference of 0.78 mm may not be significant.

In order to gain information on the impact of sub-sampling the 2D image, the similarity measure's cost as a function of a change in the transform parameters is measured by an exhaustive optimizer. This 2D surface is referred to as the 'cost function'.



(a) 1024x1024 pixels (no resize)

(b) 512x512 pixels

(c) 256x256 pixels

(d) 128x128 pixels

(e) 64x64 pixels

(f) 32x32 pixels

(g) 16x16 pixels

(h) 8x8 pixels

(i) 4x4 pixels

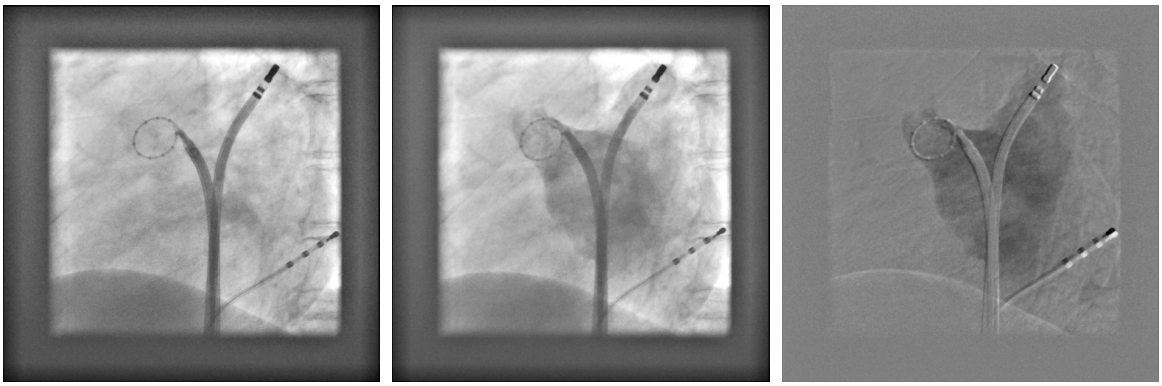
Figure 4.1 Impact of resizing P1 B on the similarity measure's cost as a function of a variation of the transform parameters. The cost is represented as a surface and the free parameters of the transforms are t_y and t_z (all others kept constant). The two background lines are the similarity measure's cost varying for one dimension while at the absolute minimum for all other dimensions.

Looking at figure 4.1 reveals that resizing down to 128x128 pixels results in a smooth cost function. From 64x64 pixels to lower resolutions, the cost function becomes increasingly erratic, with many false local minima clearly visible at the 16x16 pixels level.

Because of the result from this experiment, the registrations in the rest of this chapter are done with a 2D image resolution of 256x256 pixels.

4.1.2 Impact of Image Subtraction

Since the 2D data is a *sequence* of images showing the injection of contrast agent into the LA, it is possible to use the *variation of information between two time points*. This can be done by subtracting two images taken at a different time.



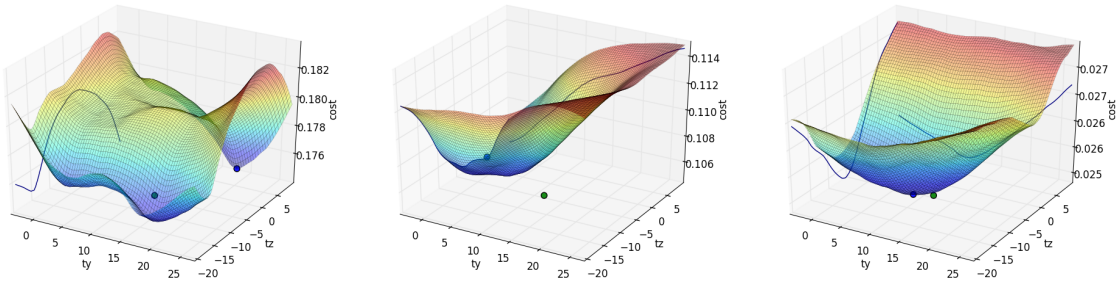
(a) P1B, frame 1

(b) P1B, frame 18

(c) P1B, frame 1 subtracted from frame 18

Figure 4.2 Subtracted and non-subtracted fluoroscopic images of P1B.

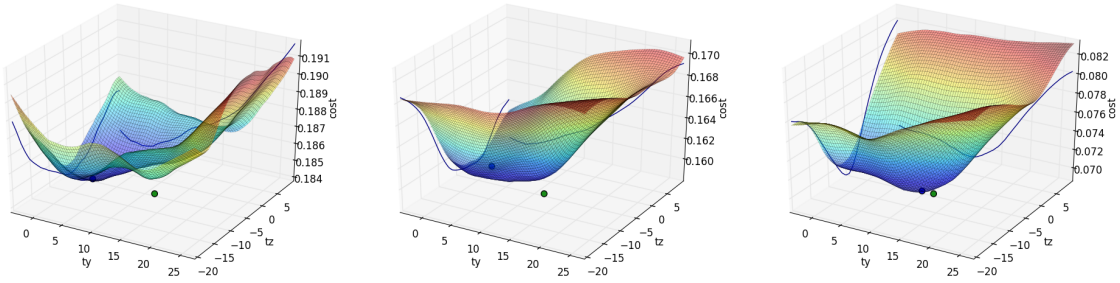
Figure 4.2 is an example of subtracted and non-subtracted images. Looking at (c) reveals that the subtraction emphasizes the LA. Also, the border, lower-left dark region as well as the patient's spine seen on the right side of (a) and (b) are all attenuated in (c).



(a) Not subtracted, MeshVerage similarity measure

(b) Not subtracted, border removed, MeshVerage similarity measure

(c) Subtractred, MeshVerage similarity measure



(d) Not subtracted, Histogram similarity measure

(e) Not subtracted, border removed, Histogram similarity measure

(f) Subtractred, Histogram similarity measure

Figure 4.3 Cost function plots for non-subtracted without border removal, non-subtracted with border removal and subtracted images. The blue dot is the minima found in the cost function and the greed dot is the ground-truth minima.

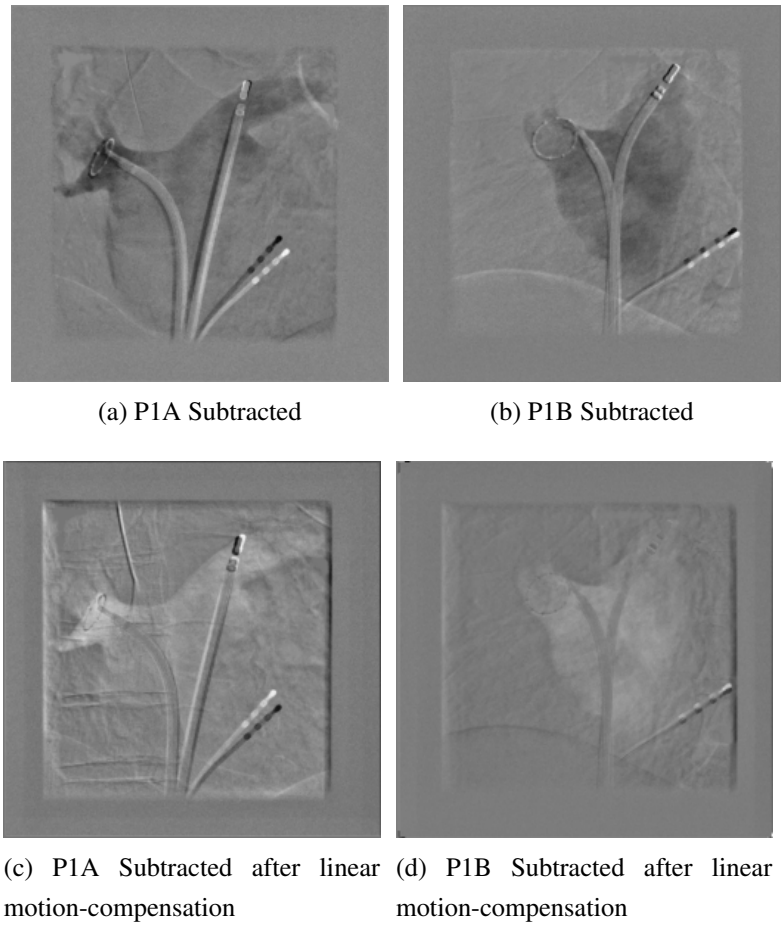
The first column of figure 4.3 shows that the similarity measures are heavily influenced by the presence of the dark border in the fluoroscopic image, especially for MeshVerage in (a). The average pixel values calculated inside and outside the mesh is impacted by the border that has an average intensity closer to the contrast agent than the rest of the image, hinting that using only averages neglects important information. Comparing (d) to (e) reveals that the Histogram similarity measure is less influenced by the presence of the border than the MeshVerage similarity measure. Nevertheless, when the border is manually removed, the similarity measures do not have a minimum concordant with the ground truth. When the images are subtracted, both similarity measures have similar graphs, and the minima are concordant with the ground truth, with the Histogram similarity measure being slightly more accurate.

The minima found by the similarity measures are not exactly equal to the ground truth. It is not due to a sampling error since the ground truth is contained in the set of transform parameters evaluated; the similarity measure simply does not have a minimum exactly at the sample place as the human operator evaluated the ground truth. The experiment show that it is better, even essential,

to use subtracted images in order to have good registration results.

4.1.3 Impact of Motion Compensation

Some of the surgical instruments' ghost artifacts are attenuated by the motion compensation, but no rigid alignment can correct for the deformation in all zones, indicating that there is significant non-rigid deformations in the images.



(a) P1A Subtracted

(b) P1B Subtracted

(c) P1A Subtracted after linear motion-compensation

(d) P1B Subtracted after linear motion-compensation

Figure 4.4 Source images for registration of case P1 with or without manually, rigid translation-only motion compensation.

Additionally, the motion compensation renders the delineation by the contrast agent less visible. This could be due to the fact that a slight misalignment of the two images to be subtracted accentuates the parts of the image where there is high-frequency noise, which correspond to the places where there is contrast agent.

Table 4.2 Registration precision for P1 with and without motion compensation, using the MeshVerage similarity measure with binary masks.

Motion-Compensation	mTRE (mm $\pm$ σmm)	mPD (mm $\pm$ σmm)
Yes	7.22 ± 1.84	7.05 ± 1.88
No	5.80 ± 0.99	5.65 ± 1.02

The results suggest that a rigid motion-compensation is not desirable as a preprocessing step.

4.2 MDIP-Based Registration for Cases of CA for AF

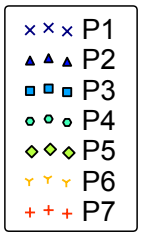
This section contains experiments using the complete registration system. The parameters for the registration are set according to the results from the previous section.

4.2.1 Performance Evaluation: Precision

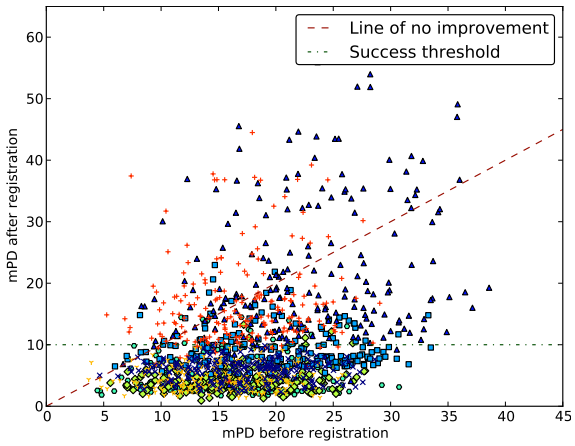
Table 4.3 Average mPD error for the 5600 registrations sorted by similarity measure and case, with variability measured using *sample* standard deviation (mPD mm $\pm$ σmm).

Case	MeshVerage	MeshVerage-DepthMask	Histogram	Histogram-DepthMask
P1	5.96 ± 1.78	4.07 ± 1.28	6.50 ± 1.07	5.49 ± 0.84
P2	22.09 ± 11.57	11.22 ± 4.06	12.96 ± 3.97	11.00 ± 4.18
P3	9.33 ± 3.19	6.83 ± 1.68	9.66 ± 2.45	7.48 ± 1.58
P4	4.91 ± 2.92	5.28 ± 1.90	4.20 ± 2.11	6.01 ± 2.11
P5	4.25 ± 1.91	10.86 ± 0.85	4.54 ± 1.66	9.24 ± 1.42
P6	4.70 ± 1.97	6.27 ± 1.79	4.40 ± 1.60	5.42 ± 0.80
P7	16.79 ± 7.36	5.63 ± 1.54	10.46 ± 2.11	5.63 ± 2.22
All Cases	9.72 ± 8.55	7.17 ± 3.32	7.53 ± 3.97	7.18 ± 2.96

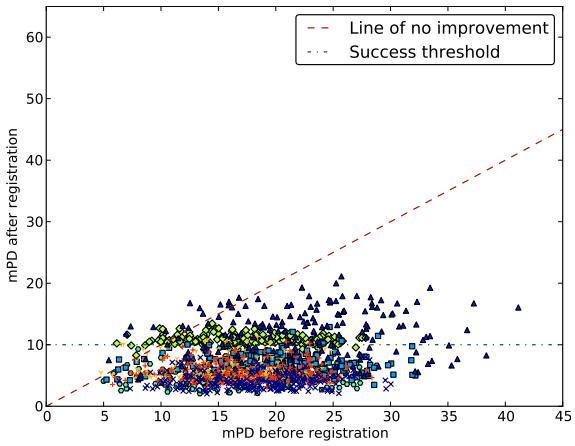
Figure 4.5 shows the registration result graphically in the manner explained in section 2.3.3.



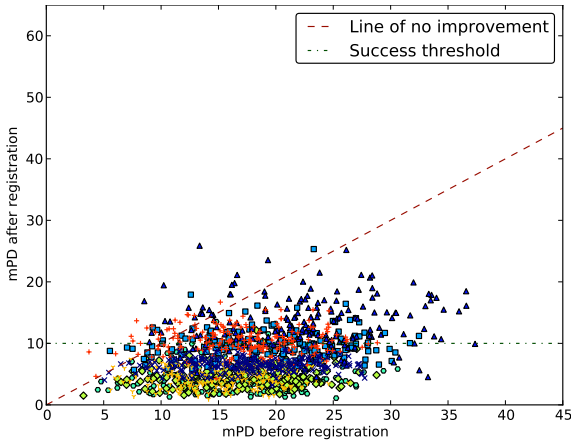
(a) Legend



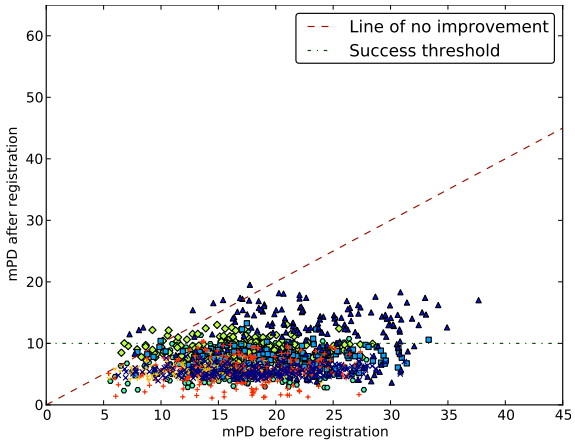
(b) MeshVerage with binary mask



(c) MeshVerage with depth-mask



(d) Histogram with binary mask



(e) Histogram with depth-mask

Figure 4.5 mPD error after registration as a function of mPD error at initialization for the different similarity measures.

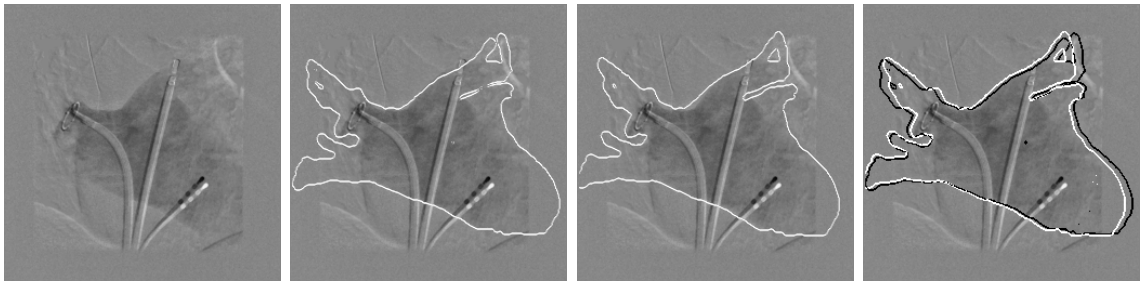
The introduction of the depth mask greatly reduces the error for the MeshVerage similarity measure. However, the Histogram similarity measure does not benefit much from using the depth-mask: the error for P7 and P3 is lowered, but is higher for P5 and P6. This indicates that the way Histogram uses the depth-masks may need to be adapted in order to profit from the additional information the depth element contribute. It is also possible that the method used to generate the

mask by exploiting the depth-values calculated by OpenGL produces artifacts that interfere with some cases. A comparison with a pure ray-casting method is needed to establish whether or not it is the case.

The error for P2 is higher for all similarity measures. The position of the ground truth for P2 is the most difficult to see: if the ground truth is difficult to position visually, the algorithm might either stumble because it is a more difficult problem or the ground truth might simply be slightly off.

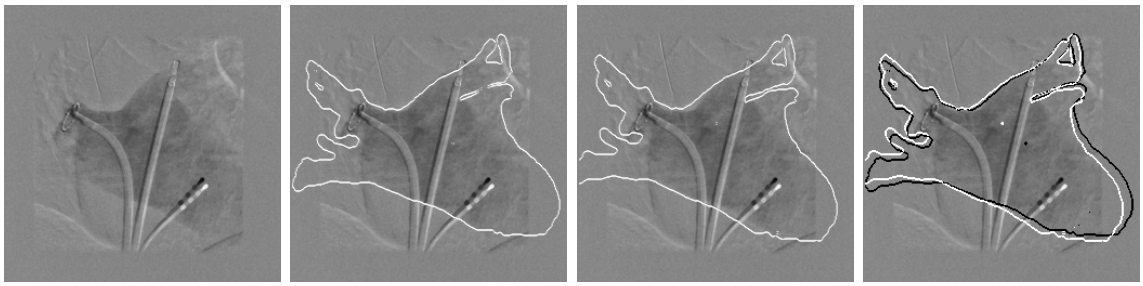
Looking at table 4.3 reveals that P5 seems to be an outlier: the use of a depth-mask doubles the error, where it significantly reduces the errors in other cases. P5's mesh was much noisier than the others, so this behavior might be due to a faulty segmentation.

The error measurement and associated figures are useful to compare the similarity measures, but does not offer a good qualitative feel for the usefulness of the registered data. The figures that follow show the visual result to expect for three levels of error (mPD): low (≈4 mm), medium (≈7 mm) and high (≈10 mm). The medium level corresponds to the average error of the best similarity measure, and the high and low levels are one standard deviation above and under the average case respectively.



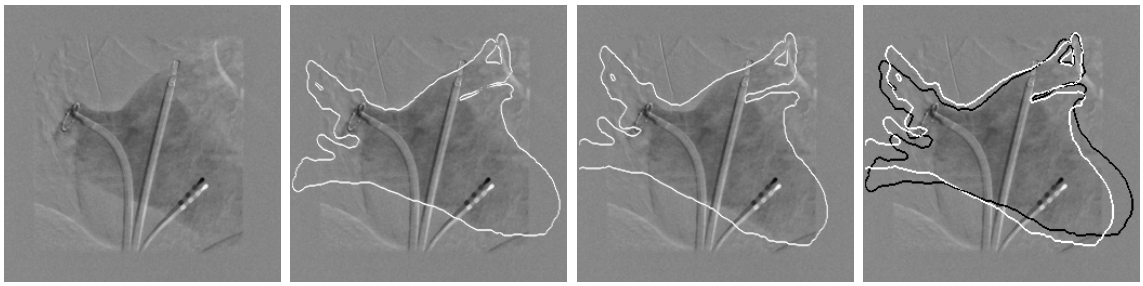
(a) X-ray (b) Ground truth (c) Registration result (d) Ground truth (black) and registration result (white)

Figure 4.6 P1A with final mPD of 4.04 mm.



(a) X-ray (b) Ground truth (c) Registration result (d) Ground truth (black) and registration result (white)

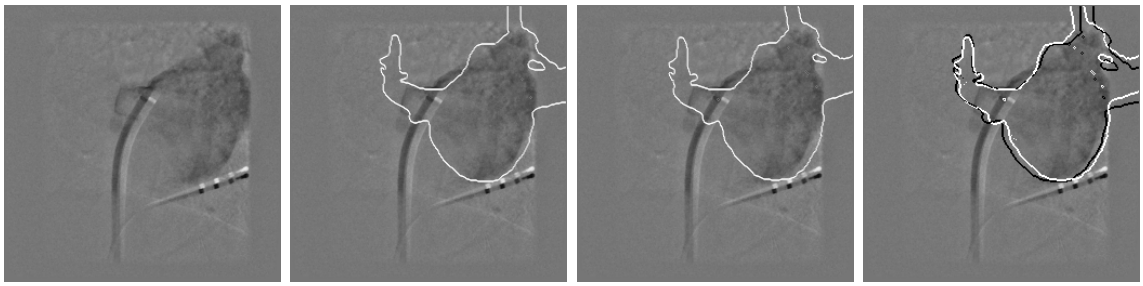
Figure 4.7 P1A with final mPD of 7.02 mm.



(a) X-ray (b) Ground truth (c) Registration result (d) Ground truth (black) and registration result (white)

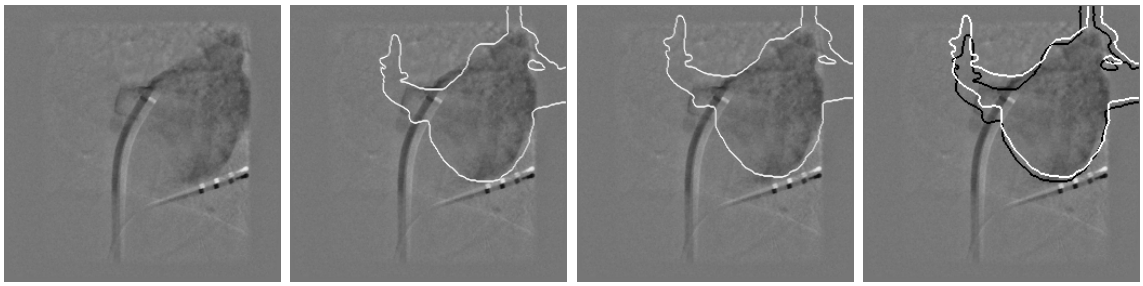
Figure 4.8 P1A with final mPD of 9.94 mm.

Notice in figures 4.6, 4.7 and 4.8 that the registration result is always at the left of the ground truth. This may be because of the lower-right image border: the ground truth is set by a human operator that bases his decision on a posteriori knowledge of the shape of the LA. The human operator can deduce that it is likely that the LA *is occluded by the border*. The registration algorithm has no information to deduce this, and is therefore pushing the mesh to the left as can be seen in figure 4.8 (c) where the lower-end of the LA fits with the delineation created by the contrast agent *and* the image border.



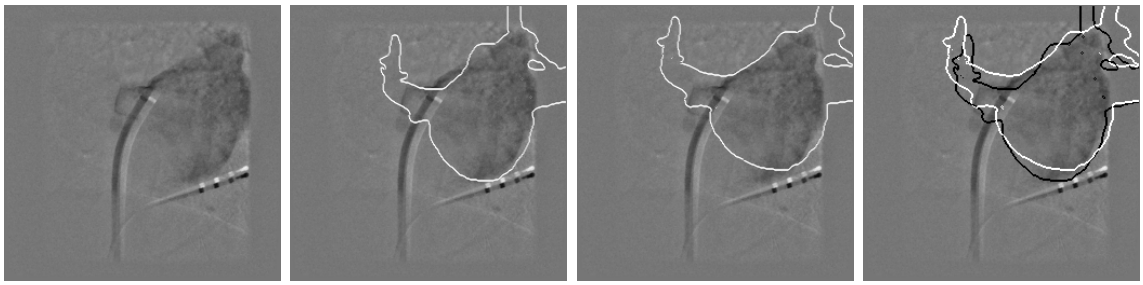
(a) X-ray (b) Ground truth (c) Registration result (d) Ground truth (black) and registration result (white)

Figure 4.9 P6B with final mPD of 4.02 mm.



(a) X-ray (b) Ground truth (c) Registration result (d) Ground truth (black) and registration result (white)

Figure 4.10 P6B with final mPD of 7.02 mm.

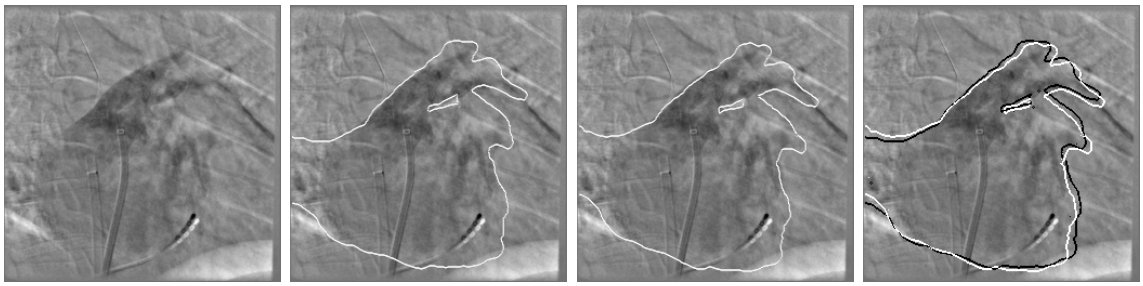


(a) X-ray (b) Ground truth (c) Registration result (d) Ground truth (black) and registration result (white)

Figure 4.11 P6B with final mPD of 10.11 mm.

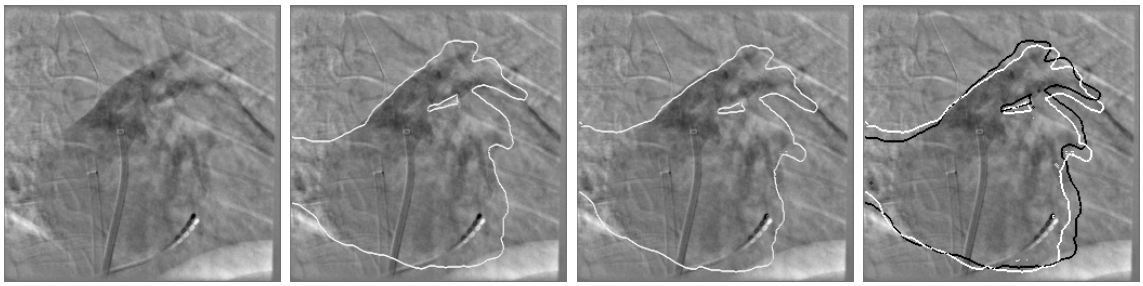
Figure 4.10 (c) delineates the LA better than the ground truth, but the pulmonary vein position is wrong. This is another case where the position of the PV is *deduced* (its position is only partially visible under fluoroscopy) using knowledge of the general shape of the LA and PVs by the human operator that defined the ground-truth. This information is not available to the registration algorithm.

Also note that the ground truths are established using biplane cases, which mean that the position for each plane is a compromise between the best position for planes A and B. This is why the ground truth is sometimes slightly wrong when looking only at one plane at a time. The fact that the two planes do not exactly match for the same 3D position is an indication that the geometry is approximate.



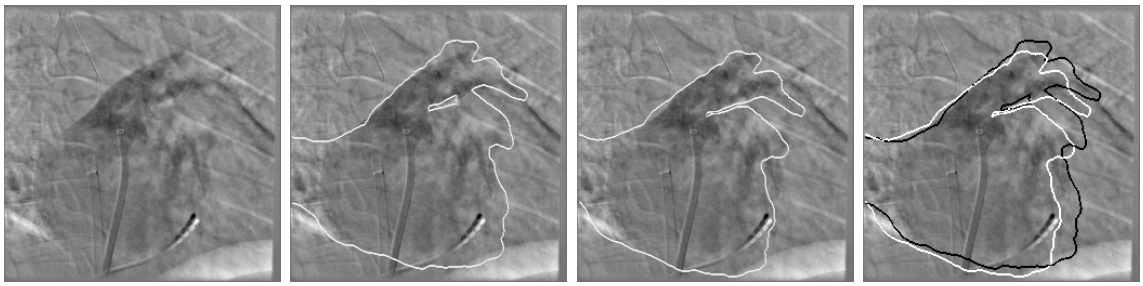
(a) X-ray (b) Ground truth (c) Registration result (d) Ground truth (black) and registration result (white)

Figure 4.12 P7A with final mPD of 4.02 mm.



(a) X-ray (b) Ground truth (c) Registration result (d) Ground truth (black) and registration result (white)

Figure 4.13 P7A with final mPD of 7.00 mm.



(a) X-ray (b) Ground truth (c) Registration result (d) Ground truth (black) and registration result (white)

Figure 4.14 P7A with final mPD of 10.02 mm.

Figures 4.12 and 4.13 illustrate the problem of the ground truth ambiguity when the delineation of the target object from the background is fuzzy such as in fluoroscopic images from cases of CA for AF. When comparing sub-figures (b) and (c), it is hard to tell which one is correct and which one is not. The registration results are *qualitatively* equivalent to the ground truth, but the mPD are 4.02 and 7.00 mm. In the case of the mPD of 7.00 mm (4.13 (c)), the registration seems a little bit worse than the ground truth when looking at the right edge, but the difference is very small. This difference is likely the lowest error threshold at which it is possible to tell that the result is worse than the ground truth. This indicates that mPDs from 0 mm to a threshold around 5 mm might be *qualitatively* as good as the ground truth.

Section 4.2.3 offers an explanation for this phenomenon: the majority of the error is contributed by zones of the PVs which are not visible under fluoroscopy therefore may routinely be positioned incorrectly by the human operation when evaluating the ground truth.

4.2.2 Performance Evaluation: Time

The speed of registration is reported as mean $\pm$ standard deviation for all cases of CA for AF. The protocol for the evaluation of this data is described in section 3.3.3.

Table 4.4 Average completion time sorted by similarity measure for biplane registrations (1400 registrations for each similarity measure).

Registration Method	Time (seconds $\pm$ σ)
MeshVerage	19.03 ± 9.38
MeshVerage DepthMask	31.00 ± 14.64
Histogram	127.00 ± 61.55
Histogram DepthMask	191.81 ± 97.36

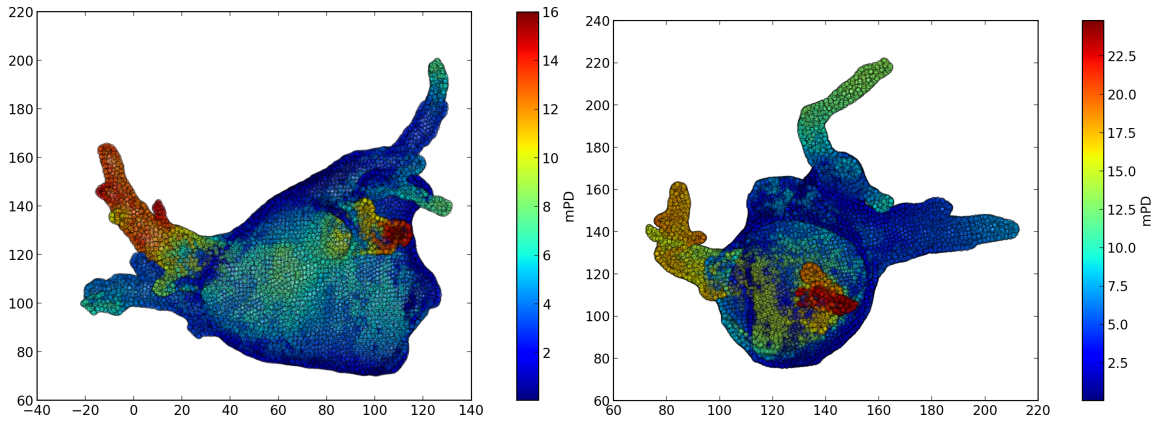
The closest comparison to our registration solution is work done by Knecht et al. [9]; their team provides registration in cases of CA for AF based on the segmentation of the patient's spine from the CT image which is then registered with the real-time fluoroscopic images. They report that their registration takes 420 ± 120 seconds, about 13.5 time as long as the 'MeshVerage DepthMask' method we introduce. For spine registration, Van De Kratts et al. [36] report a registration time of 25 seconds for a gradient-based application and 9 minutes for an intensity-based registration.

The time that we report in table 4.4 is from non-optimized C++ code running on a low-end laptop in single-threaded mode. It is hard to have a fair comparison with Knecht et al. since they do not mention implementation details such as the hardware used and if it is parallelized or not. Also, our implementation uses biplane registration — in monoplane cases it is likely that it would take half as much time to complete. In our implementation, most of the time is spent in getting a cost for the best-neighbors in the optimization process, which requires 24 evaluations of the similarity measure per iteration (12 per plane). This is completely parallelizable: a multithreaded implementation of the algorithm running on an N-core machine is expected to approach a speedup factor of N, for factors up to 24.

MDIP-based registration therefore significantly outperforms other projection-based algorithms in terms of speed, and may outperform the gradient-based approaches once parallelized.

4.2.3 Error Distribution as a Function of the Position on the Mesh

Since the error is computed on every vertex of the mesh, it is possible to display the distribution of the error depending on the local position on the mesh.



(a) Plane A, mPD of 4.95 mm.

(b) Plane B, mPD of 8.56 mm.

Figure 4.15 P6 error distribution on the mesh. The mPD for this case is 6.75 mm.

Figure 4.15 clearly shows that the error is higher on the PVs than on the LA. Since the ablation targets are located on the LA *around* the PVs [2] (see figure 2), the *application error* is much lower than the reported mPD when calculated on the whole mesh (including the PVs). For the case displayed by figure 4.15, the application error (mPD on the LA around the PVs) is estimated to be around 2 to 3 mm, while the mPD for the whole mesh is 6.75 mm.

This indicates that the error measured for the MDIP-based algorithm would probably be much lower if it was measured using manually-selected fiducial points on the LA rather than on the PVs, as some other papers do.

This also puts in context preliminary results where we reported an mPD of 6.30 ± 2.55 mm using the 'MeshVerage' similarity measure with binary masks and meshes *where the pulmonary veins were manually cut-off* [59]. Supposing that using a depth-mask is equivalent to using a cut mesh, the mPD *measured* is lower when the meshes are cut for the same surgical cases: discarding the PVs may significantly lower the error *measured*, without lowering the actual and application error. Therefore, a comparison of the error levels measured reveals that it is probably better to use depth-masks or the Histogram similarity measure rather than manually cutting the PVs prior to registration.

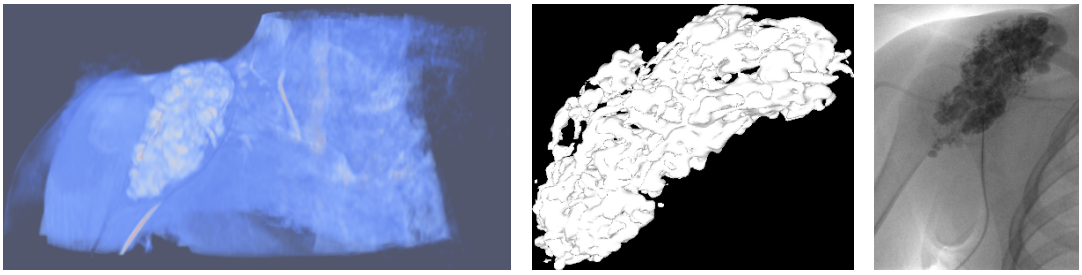
4.3 MDIP-Based Registration for Cases of SfVM

This section contains the result of registrations for cases of sclerotherapies for venous malformation (SfVM). This is an attempt to see if the same methodology used in CA for AF cases can successfully be applied to other type of surgeries. The difficulty in cases of SfVM is that the malformations can be found anywhere on the patient's body [45], making it difficult to establish a registration protocol as well as validation scheme. The cases that we have available are monoplane,

which makes it arduous to establish a ground-truth because of the lack of information in the out-of-plane direction. The meshes in cases of SfVM were obtained by visual segmentation from the MR image using the ParaView program [63]. The validation for the cases of SfVM is done visually, with a case-study style that highlights the challenges typical to this type of surgery. The section ends with recommendation on further development required in order to improve the performance of the MDIP-based algorithm for SfVM.

4.3.1 Case Study: Shoulder

Venous malformations can be present on any part of the body, which make it difficult to establish a registration protocol. The zones of the image that are exploitable for registration vary from one case to another. One approach to the problem is to exploit information that is constant in most VM case: the presence of an embolized venous malformation that is easily seen on both fluoroscopy and MRI.



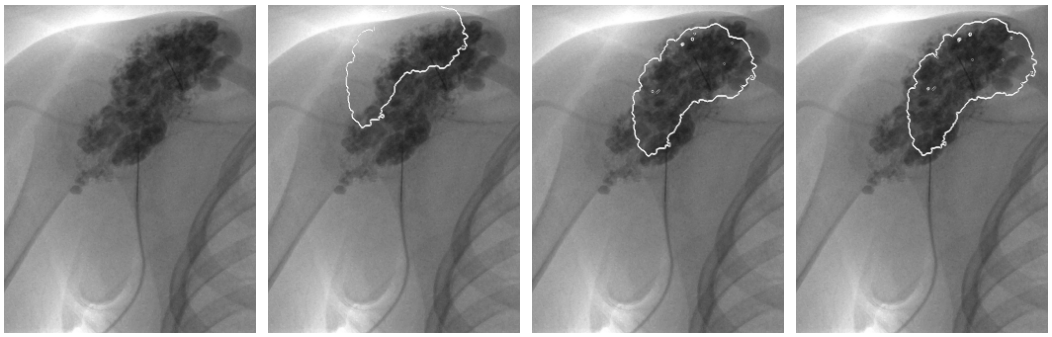
(a) MRI

(b) Mesh segmented from MRI

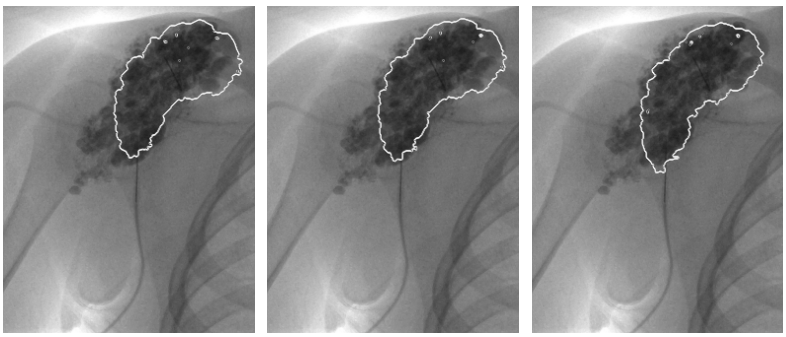
(c) Fluoroscopic image

Figure 4.16 Case of SfVM on a shoulder shown under different modalities.

In this case, the registration uses the Histogram similarity measure with a binary mask. The parameters are exactly the same as used in the CA for AF cases.



(a) Fluoroscopic image (b) Registration initial pose (c) Registration iteration 5 (d) Registration iteration 10



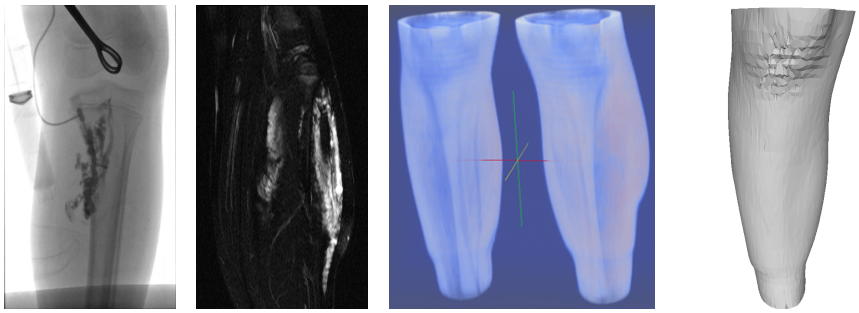
(e) Registration iteration 15 (f) Registration iteration 20 (g) Registration iteration 46 (final)

Figure 4.17 Registration result for the shoulder case using the VM to drive the registration.

Using the embolization as the feature to register has the drawback that it might not work well when the venous malformation is relatively shapeless or that the MRI does not match the fluoroscopy.

4.3.2 Case Study: Leg

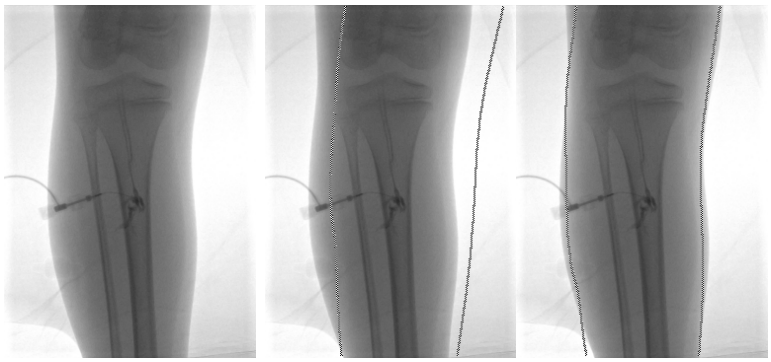
In the ‘leg case’, the intraoperative fluoroscopic images show the full organ from an outside view. The outside area has a distribution of pixel intensities that is drastically different from the organ, which facilitates the registration. The protocol is the same as was done for cases of CA for AF: the MRI is segmented into a mesh, and then the mesh is registered to the intraoperative fluoroscopic images.



(a) Fluoroscopic image (b) MRI slice (c) MRI 3D view (d) Mesh extracted from the MRI

Figure 4.18 Case of SfVM on a leg shown under different modalities.

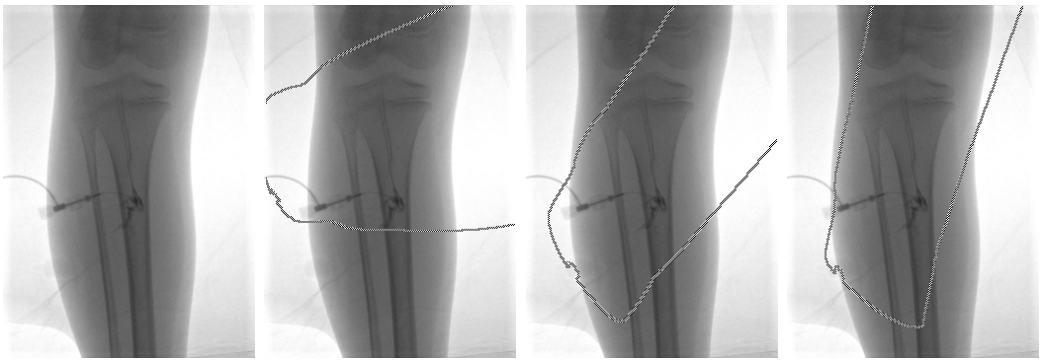
For the registration reported below, the Histogram similarity measure with binary mask is used. Tests were also conducted with the 'MeshVerage' similarity measure, and the performance was almost identical.



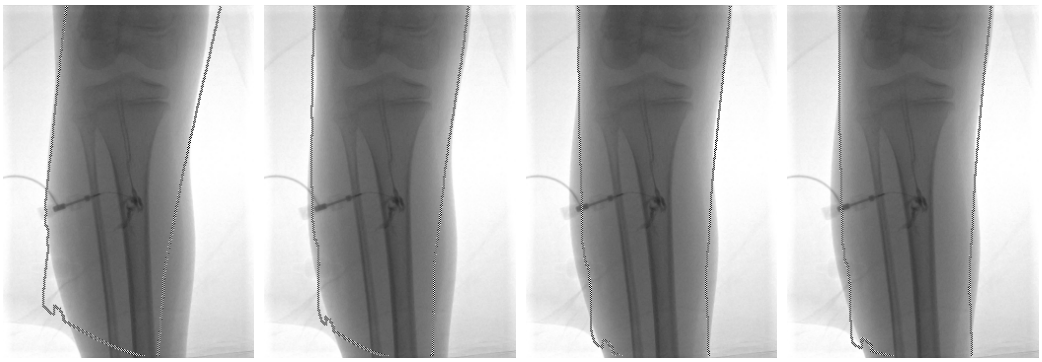
(a) Fluoroscopic image (b) Initilization of registration. (c) Registration result

Figure 4.19 Registration of a leg under monoplane fluoroscopy using the Histogram similarity measure and a binary mask.

The mesh is properly positioned at the end of the registration and clearly overlaps with the leg's border. However, since the leg is relatively shapeless in the rotation around its principal axis, it is possible that there is a rotational error that cannot be identified by merely looking at the projective overlap with the target organ. This case also registers well even when the initial guess is extremely off. Figure 4.20 shows a registration with such an initial position.



(a) Fluoroscopic image (b) Initilization of registration (c) Registration iteration 10 (d) Registration iteration 20

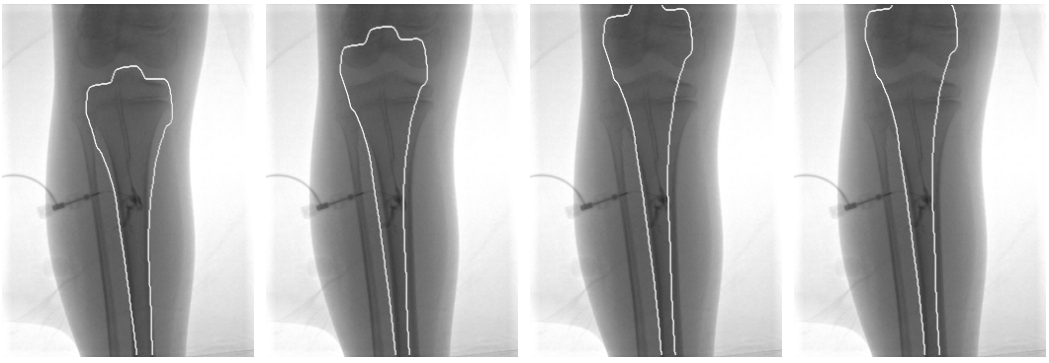


(e) Registration iteration 30 (f) Registration iteration 40 (g) Registration iteration 50 (h) Registration iteration 62 (final)

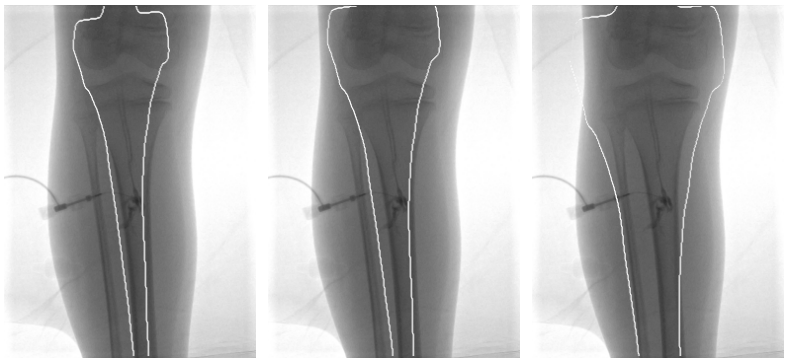
Figure 4.20 Same registration as figure 4.19, but with a starting point that is an extreme corruption as initial position. The registration algorithm is nevertheless able to recover a reasonable pose.

The registration overlap is correct, but the end result seems a bit slanted. However, it is expected that the result is slightly worse than for a realistic case such as seen in figure 4.19 because of the extreme initial error.

Using the inside/outside border of the organs seems to give a reasonable approximation of the registration, albeit without aligning the images perfectly. In many cases of SfVM, there are prominent bones which have a lot of fine details; they can also easily be segmented from the MR image. Since these fine details might be helpful to gain registration accuracy, we conduct an experiment to register the same case but with the tibia as the feature.



(a) Initilization of registration (b) Registration iteration 5 (c) Registration iteration 10 (d) Registration iteration 15

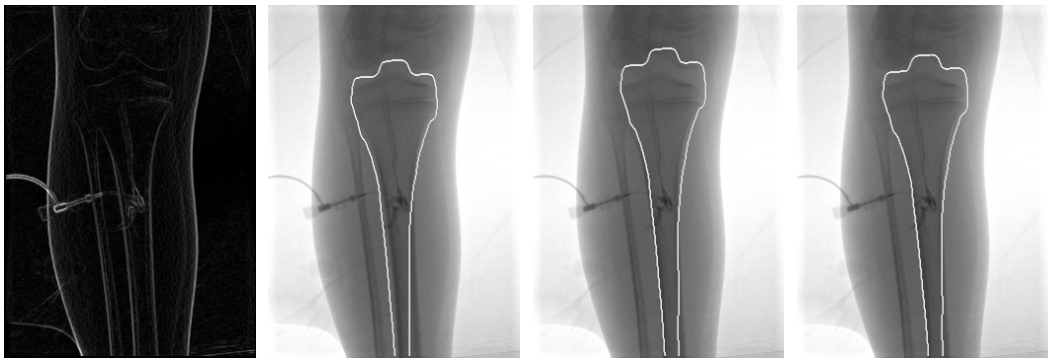


(e) Registration iteration 20 (f) Registration iteration 40 (g) Registration iteration 84 (final)

Figure 4.21 Registration of the tibia using the Histogram similarity measure.

The registration from figure 4.21 illustrates a limitation of the MDIP-based registration when the similarity measure uses statistical measures from the whole image. The registration does successfully delineate the bones, but it in an extreme way that encompasses all the bony structures inside the leg. The algorithm was originally developed to solve cases where the delineation of the organ is fuzzy, where it is a good thing to bundle together the parts of the image that are statistically similar. In the cases where one bone must be isolated from the others, such a behavior is harmful. Also, there are three zones in the images where the pixels have markedly different statistical properties: the inside of the bones, the inside of the leg (excluding the bones), and the outside of the leg. Doing the partition using only two groups (as is adequate for cases of CA for AF) is probably not advisable in this case.

Next is an example of the same registration, but based on a simple similarity measure that maximizes the 3D-2D gradients obtained with a Sobel filter. The idea is show that a simple modification of the MDIP-based algorithm successfully uses the gradients to register bony structures.



(a) Fluoroscopic image after Sobel filtering (b) Initilization of registration (c) Registration iteration 5 (d) Registration iteration 9 (final)

Figure 4.22 A registration of the tibia based on 3D-2D gradients. Note that the left and right lower parts of the tibia are properly delineated by the mesh, but the top part stops at the next bone, slightly misplacing the upper end of the tibia.

The registration algorithm used for figure 4.22 has extremely limited range: it cannot correct for high initial pose errors. However, it demonstrates that the similarity measures have to consider statistical elements other than direct pixel intensities. It remains an open question on how to best integrate the use of gradients in an MDIP-based algorithm (the example shown above is a simple maximization of the value of the gradient along the mesh's border, similar to what is done in [55]). An hybrid algorithm that first finds a crude pose based on the inside/outside border of an organ or an embolization then switches to gradient matching for bony structure in order to get a precise pose is advised. However, such a method requires that many elements be present in the same case, which does not allow solving all SfVM cases using the same protocol.

4.4 General Discussion

This section reports the results of the application of the MDIP-based registration algorithm on cases of CA of AF, for which an in-depth numerical analysis as well as visual results have been presented. Visual and case-study type analysis were done for the SfVM cases, demonstrating that MDIP-based registration is extendable to other type of surgeries, but at the same time highlighting that modifications to the similarity measure depending on the specificities of the surgical case is necessary for this extension.

There is one aspect that has not yet been discussed in the error analysis: the factors contributing to misregistration, which come from two distinct sources: the algorithm itself, and the experiment setup. The rest of this section reviews these elements and ends with a discussion of the relevance

of the error analysis.

4.4.1 Errors Due to the MDIP-Based Algorithm

There are two aspects of the MDIP-based algorithm that may contribute to misregistration and warrant further analysis: the depth-mask creation process as well as the selection of the frames to subtract in the sequence showing the injection of the contrast agent in the LA. But before talking about what contributes to the error, we comment on an easy avenue to gain precision that was intentionally avoided.

The *optimizer* used for the MDIP-based algorithm was willfully not selected to get the minimum possible error: the techniques laid-out in section 2.1.4, which are proven ways to lower the average error for registration algorithms, were not adopted. This decision is motivated by a desire to be sensible to subtle variations in the algorithm's behavior — such optimizations make it harder to measure the impact of parameters change as these clever optimization techniques may cloak underlying weaknesses. As the project evolves from the validation phase to a clinical trial, these methods can now be implemented in order to lower the expected error. Fortunately, the MDIP-based algorithm is perfectly compatible with these tentative changes as they usually involve modification of the optimizer, not the similarity measures (which are more strongly tied to the MDIP-based algorithm).

The *depth-mask process* described in section 3.1.1 drastically improves the average registration for the MeshVerage similarity measure, but it does not have the same dramatic positive impact for the Histogram similarity measure. The parameters used in the creation of the depth-mask influence the thickness of the PVs for which the similarity measure assumes that the 3D image will not match the fluoroscopic image. This information is not integrated in the same manner for the different similarity measures; it should therefore not be assumed that the same settings influence all the similarity measures uniformly. Optimizing these settings for each similarity measure might enhance their performance. Also, the current generation of depth-masks hinges on a clever approximation of the mesh's local depth that exploits calculations done internally by the graphics card in order to avoid the computation of the distance-transform. However, the impact on precision of that approximation remains to be evaluated. A distance-transform based projection method should be compared to our current approximation in order to verify if we sacrifice quality for speed in this process.

The *manual selection of two frames to subtract* from the injection of contrast agent video sequence is the last manual remnant in our otherwise completely automatic registration algorithm. This selection is much quicker than doing a complete manual registration – our system is an improvement

over a completely manual registration – but it is an operation that needs to be discarded before we can claim that our process is completely automatic. There are a number of ways this can be accomplished. If the beginning and end of the injection of contrast agent are signaled by operative hardware, it is possible to systematically select one frame in the middle of the contrast injection sequence. It is also possible to average a few images during the injection, and to subtract this averaged image with the first image. Another solution is to detect the zones with contrast agent, and to use that information for frame selection or directly for registration. Note that, because of pulmonary and patient movement, the registration result is likely to be different depending on the frame(s) selected. No matter what the precision of the registration on a single frame is, there is a limited clinical utility to be extremely precise since the exact position of the LA moves during the cardiac cycle and patient respiration. The next section (4.4.2) discusses other aspects that limit the precision. It concludes with a proposition that when the precision is inherently limited, the robustness of the algorithm should be the principal metric considered to drive research effort, not the absolute, single-frame precision.

4.4.2 Errors Due to the Experiment Setup

There are three sources of error that are external to the registration algorithm: the approximate geometry, the quality of the fluoroscopic images and the precision of the ground truth. After covering the cause of errors due to the algorithm and the experiment setup, we discuss whether the precision is the right metric to drive research effort.

The *geometry* is obtained through manual refinement of a crude initial approximation. Constraints on the geometry are assumed in order to make this process feasible, some which may be inaccurate (see section 3.2.2): it is unknown how much this degrades the precision. Since the geometry is available during the surgeries, allowing a serialization of the geometry to a file which is later sent along with the medical images for analysis can correct this error source. This is currently being implemented and future cases will have the full geometry defined. It is worth indicating that our algorithm is expected to behave better if used clinically, where the geometry is correctly defined.

The *quality of the fluoroscopic images* taken during the injection of the contrast agent has an impact on the quality of the registration that has not been evaluated. Table 4.3 and figure 4.5 reveal that the average registration precision varies widely depending on the case. P2 is by far the case that has the worse registration error for all similarity measures used, and not surprisingly it is the case where visually it is also the most difficult to establish the ground-truth. The fluoroscopic data comes from operation not targeted to the application of automatic registration algorithms. Therefore, if the injection of contrast agent is done with more care to completely delineate the whole LA

in one single injection, the average registration result might be better. In addition, we did not systematically use frames at the same position in the cardiac cycle (simply because we did not have access to the electrocardiogram data). Synchronization with the cardiac cycle may allow increasing the consistency of the results.

The *precision of the ground truths* needs to be evaluated. The truth-values used as a comparison basis for our algorithm is defined by a human operator, who registers with an inherently limited precision. It is necessary to evaluate the consistency of the registration from one operator to another as well as for one operator at different times. For example, if for exactly the same surgical case, an operator registers with a variability of 5 mm (mPD), it means that the ground truths cannot be used to evaluate an algorithm's precision for under that 5 mm threshold. This evaluation should be part of a clinical trial, which is described in section 5.0.3.

We spent a great deal of effort tuning the algorithm to register with high-precision, but there is a point where is it worth asking if the precision we measure is really the most important metric to optimize in order to deliver a clinically effective algorithm. Since the border of the organ is fuzzy, likely to move around due to respiratory and cardiac motion and dependent on the frame used from the contrast agent injection sequence, it may be of little added value to spend a lot of effort to reach a very high precision for a single point in time. The really important aspect for the registration is to fail infrequently, rather than have a millimeter-precision at one point in time. The robustness of the algorithm —- the extent to which the algorithm never fails disastrously — might therefore be more important than high-precision. As can be seen in figure 4.5 of section 4.2.1, the algorithm is very robust to large corruptions of the initial pose and rarely degrades the pose when using the best performing similarity measures. This is a strong point in favor of the MDIP-based algorithms.

CHAPTER 5 : CONCLUSION

The principal objective of this research project was to conceive a registration algorithm that automatically aligns a preoperative MR image to intraoperative biplane X-ray fluoroscopy images for cases of CA for AF. There were two significant challenges to overcome. First, current solutions take several minutes to complete, which is cumbersome when the registration needs to be updated intraoperatively because of patient movement. Second, multimodal MR/X-ray fluoroscopy registration is a notoriously difficult problem, further complicated in cases of CA for AF because the 3D and 2D images do not match at the level of the PVs during the injection of contrast agent.

We introduced the novel MDIP registration algorithm which uses a mesh segmented from the MR image to partition the intraoperative fluoroscopic images. The discrepancies of the pixels' statistical properties contained in the partitions are used by a similarity measure to infer a cost, which drives the registration. This addresses the difficult MR/X-ray registration problem by transforming a 3D-2D intermodal registration into a 2D-2D intramodal problem. Also, the partition can be done very quickly, which enables fast registrations. In addition, we developed a modification to the similarity measures to account for the partial or absent match of the PVs with the concept of statistical group-affiliation based on depth-mask intensity.

Our research hypothesis that this novel registration algorithm is faster than current methods is successfully validated by experiments on 7 clinical cases of CA for AF. Our solution completes in about 30 seconds for biplane registration, much faster than the volumetric projection-based solutions which typically require several minutes. We validated the MDIP-based algorithm on 5600 registrations covering 7 cases and four similarity measures by reporting the mPD error at the beginning and end of the registrations. Visual examples of typical registration results as well as one standard deviation above and under the mean error are also produced.

Aside from potentially providing registration that is clinically practical for cases of CA for AF, the MDIP-based algorithm is generalizable to other types of 3D-2D registration problems such as SfVM. It is especially promising when the 3D modality is MR images because it has the potential to reduce this class of notoriously difficult registration problem to a less elaborate segmentation problem that has a straightforward solution.

5.0.3 Recommendations and Future Work

We make an informed claim that the MDIP-based registration algorithm introduced in this thesis is adequate for use in cases of CA for AF. However, the ultimate evaluation of our system is whether or not it can outperform a human operator in terms of speed and precision. To verify this, we propose a two phases clinical trial. The first phase consists of recording the fluoroscopic video feed as well as the manual registration of the mesh overlay during surgeries. This recorded data can then be used to compare the results of the registration done manually with the result from our algorithm. This simulation mode should also allow doing the registrations manually from the recorded surgeries in order to evaluate the level of discrepancy in the registration done by different medical specialists and the same medical specialist at different times. If the results from our algorithm compares favorably to the manual results, then the second-stage full clinical trial using the registration algorithm to replace manual registration can be put in place.

The MDIP-based registration algorithm introduced in this thesis is not limited to cases of CA for AF. In order to evaluate how portable to other type of surgeries it is, we demonstrate that the algorithm is successful for cases of SfVM. Since the SfVM cases are monoplane, this cues that the algorithm is not limited to biplane imaging. However, the difficulties encountered on some of the cases indicate that new MDIP-based similarity measures needs to be tuned in order to enhance the breadth of surgical cases that the MDIP-based algorithm can handle. We demonstrate that using a hybrid approach that incorporates gradient information in the similarity measure can improve the registration for some of the difficult cases where the structure to register contains bones.

REFERENCES

[1] R. Galloway and T. Peters, "Overview and history of image-guided interventions," in *Image-Guided Interventions* (T. Peters and K. Cleary, eds.), pp. 1–21, Springer US, 2008.

[2] A. Ames and W. G. Stevenson, "Catheter ablation of atrial fibrillation," *Circulation*, vol. 113, no. 13, pp. e666–e668, 2006.

[3] J. Sra, G. Narayan, D. Krum, A. Malloy, R. Cooley, A. Bhatia, A. Dhala, Z. Blanck, V. Nangia, and M. Akhtar, "Computed tomography-fluoroscopy image integration-guided catheter ablation of atrial fibrillation," *Journal of Cardiovascular Electrophysiology*, vol. 18, pp. 409–414, 2007.

[4] R. Razavi, D. L. Hill, S. F. Keevil, M. E. Miquel, V. Muthurangu, S. Hegde, K. Rhode, M. Barnett, J. van Vaals, D. J. Hawkes, and E. Baker, "Cardiac catheterisation guided by mri in children and adults with congenital heart disease," *The Lancet*, vol. 362, no. 9399, pp. 1877 – 1882, 2003.

[5] P. M. Kistler, K. Rajappan, M. Jahngir, M. J. Earley, S. Harris, D. Abrams, D. Gupta, R. Liew, S. Ellis, S. C. Sporton, and R. J. Schilling, "The impact of ct image integration into an electroanatomic mapping system on clinical outcomes of catheter ablation of atrial fibrillation," *Journal of Cardiovascular Electrophysiology*, vol. 17, no. 10, pp. 1093–1101, 2006.

[6] D. Bhakta and J. M. Miller, "Principles of electroanatomic mapping," 2008.

[7] T. S. Fahmy, H. Mlcochova, O. M. Wazni, D. Patel, R. Cihak, M. Kanj, S. Beheiry, J. D. Burkhardt, T. Dresing, S. Hao, P. Tchou, J. Kautzner, R. A. Schweikert, M. Arruda, W. Saliba, and A. Natale, "Intracardiac echo-guided image integration: Optimizing strategies for registration," *Journal of Cardiovascular Electrophysiology*, vol. 18, no. 3, pp. 276–282, 2007.

[8] J. Sra, D. Krum, A. Malloy, M. Vass, B. Belanger, E. Soubelet, R. Vaillant, and M. Akhtar, "Registration of three-dimensional left atrial computed tomographic images with projection images obtained using fluoroscopy," *Circulation*, vol. 112, no. 24, pp. 3763–3768, 2005.

[9] S. Knecht, H. Skali, M. D. O'Neill, M. Wright, S. Matsuo, G. M. Chaudhry, C. I. Haffajee, I. Nault, G. H. Gijsbers, F. Sacher, F. Laurent, M. Montaudon, O. Corneloup, M. Hocini, M. Haïssaguerre, M. V. Orlov, and P. Jaïs, "Computed tomography–fluoroscopy overlay evaluation during catheter ablation of left atrial arrhythmia," *Europace*, vol. 10, no. 8, pp. 931–938, 2008.

[10] J. Sra, "Cardiac image integration implications for atrial fibrillation ablation," *Journal of Interventional Cardiac Electrophysiology*, vol. 22, pp. 145–154, 2008.

[11] J. Dubois, G. Soulez, V. L. Oliva, M.-J. Berthiaume, C. Lapierre, and E. Therasse, "Soft-tissue venous malformations in adult patients: Imaging and therapeutic issues1," *Radiographics*, vol. 21, no. 6, pp. 1519–1531, 2001.

[12] A. A. de Lorimier, "Sclerotherapy for venous malformations," *Journal of Pediatric Surgery*, vol. 30, no. 2, pp. 188 – 194, 1995.

[13] A. Webb, "Introduction to Biomedical Imaging," *Medical Physics*, p. 2267, 2003.

[14] K. S. Rhode, M. Sermesant, D. Brogan, S. Hegde, J. Hipwell, P. Lambiase, E. Rosenthal, C. Bucknall, S. A. Qureshi, J. S. Gill, R. Razavi, and D. L. G. Hill, "A system for real-time xmr guided cardiovascular intervention," *Medical Imaging, IEEE Transactions on*, vol. 24, no. 11, pp. 1428–1440, 2005.

[15] D. J. Brenner and E. J. Hall, "Computed tomography — an increasing source of radiation exposure," *New England Journal of Medicine*, vol. 357, no. 22, pp. 2277–2284, 2007.

[16] P. Markelj, D. Tomaževic, B. Likar, and F. Pernuš, "A review of 3d/2d registration methods for image-guided interventions," *Medical Image Analysis*, no. 0, pp. –, 2010.

[17] E. Trucco and A. Verri, *Introductory Techniques for 3-D Computer Vision*. Upper Saddle River, NJ, USA: Prentice Hall PTR, 1998.

[18] J. Yang, Y. Wang, Y. Liu, S. Tang, and W. Chen, "Novel approach for 3-d reconstruction of coronary arteries from two uncalibrated angiographic images," *Image Processing, IEEE Transactions on*, vol. 18, pp. 1563 –1572, july 2009.

[19] S. Miao, R. Liao, and Y. Zheng, "A hybrid method for 2-d/3-d registration between 3-d volumes and 2-d angiography for transcatheter aortic valve implantation (tavi)," in *Biomedical Imaging: From Nano to Macro, 2011 IEEE International Symposium on*, pp. 1215 –1218, 2011.

[20] A. Khamene, P. Bloch, W. Wein, M. Svatos, and F. Sauer, "Automatic registration of portal images and volumetric ct for patient positioning in radiation therapy," *Medical Image Analysis*, vol. 10, no. 1, pp. 96 – 112, 2006.

[21] C. Studholme, D. Hill, and D. Hawkes, "An overlap invariant entropy measure of 3d medical image alignment," *Pattern Recognition*, vol. 32, no. 1, pp. 71 – 86, 1999.

[22] M. Jenkinson and S. Smith, "A global optimisation method for robust affine registration of brain images," *Medical Image Analysis*, vol. 5, no. 2, pp. 143 – 156, 2001.

[23] C. D. Manning, P. Raghavan, and H. Schtze, *Introduction to Information Retrieval*. New York, NY, USA: Cambridge University Press, 2008.

[24] R. O. Duda, P. E. Hart, and D. G. Stork, *Pattern Classification*. Wiley-Interscience, 2 ed., 2001.

[25] M. Powell, "A fast algorithm for nonlinearly constrained optimization calculations," in *Numerical Analysis* (G. Watson, ed.), vol. 630 of *Lecture Notes in Mathematics*, pp. 144–157, Springer Berlin / Heidelberg, 1978.

[26] W. H. Press, S. A. Teukolsky, W. T. Vetterling, and B. P. Flannery, *Numerical Recipes: The Art of Scientific Computing*. New York, NY, USA: Cambridge University Press, 3 ed., 2007.

[27] J. A. Nelder and R. Mead, "A simplex method for function minimization," *The Computer Journal*, vol. 7, no. 4, pp. 308–313, 1965.

[28] H. Sundar, A. Khamene, C. Xu, F. Sauer, and C. Davatzikos, "A novel 2d-3d registration algorithm for aligning fluoro images with 3d pre-op ct/mr images," in *Proceedings of the Society of Photo-Optical Instrumentation Engineers (SPIE)* (K. R. Cleary, R. L. Galloway, and Jr, eds.), vol. 6141, pp. 760–766, SPIE.

[29] M. Mahfouz, W. Hoff, R. Komistek, and D. Dennis, "A robust method for registration of three-dimensional knee implant models to two-dimensional fluoroscopy images," *Medical Imaging, IEEE Transactions on*, vol. 22, pp. 1561 –1574, dec. 2003.

[30] D. Gorse, A. Shepherd, and J. Taylor, "A classical algorithm for avoiding local minima," in *Proceedings of the World Congress on Neural Networks*, pp. 364–369, 1994.

[31] G.-A. Turgeon, "2D-3D registration of coronary angiograms for cardiac procedure planning and guidance," *Medical Physics*, vol. 32, p. 3737, 2005.

[32] J. Hipwell, G. Penney, R. McLaughlin, K. Rhode, P. Summers, T. Cox, J. Byrne, J. Noble, and D. Hawkes, "Intensity-based 2-d - 3-d registration of cerebral angiograms," *Medical Imaging, IEEE Transactions on*, vol. 22, pp. 1417 –1426, nov. 2003.

[33] G. Penney, J. Weese, J. Little, P. Desmedt, D. Hill, and D. hawkes, "A comparison of similarity measures for use in 2-d-3-d medical image registration," *Medical Imaging, IEEE Transactions on*, vol. 17, pp. 586 –595, aug. 1998.

[34] F.-F. Yin, Q. Gao, H. Xie, D. F. Nelson, Y. Yu, W. Kwok, S. Totterman, M. C. Schell, and P. Rubin, "Mr image-guided portal verification for brain treatment field," *International Journal of Radiation Oncology, Biology, Physics*, vol. 40, no. 3, pp. 703 – 711, 1998.

[35] P. Markelj, D. Tomazevic, F. Pernus, and B. Likar, "Robust gradient-based 3-d/2-d registration of ct and mr to x-ray images," *Medical Imaging, IEEE Transactions on*, vol. 27, pp. 1704 – 1714, dec. 2008.

[36] E. van de Kraats, G. Penney, D. Tomazevic, T. van Walsum, and W. Niessen, "Standardized evaluation methodology for 2-d-3-d registration," *Medical Imaging, IEEE Transactions on*, vol. 24, no. 9, pp. 1177 –1189, 2005.

[37] T. Rohlfing and C. R. Maurer, Jr., "A novel image similarity measure for registration of 3-d mr images and x-ray projection images," in *Proceedings of the 5th International Conference on Medical Image Computing and Computer-Assisted Intervention-Part II*, MICCAI '02, (London, UK, UK), pp. 469–476, Springer-Verlag, 2002.

[38] R. McLaughlin, J. Hipwell, D. Hawkes, J. Noble, J. Byrne, and T. Cox, "A comparison of a similarity-based and a feature-based 2-d-3-d registration method for neurointerventional use," *Medical Imaging, IEEE Transactions on*, vol. 24, pp. 1058 –1066, aug. 2005.

[39] S. Miao, R. Liao, and M. Pfister, "Toward smart utilization of two x-ray images for 2-d/3-d registration applied to abdominal aortic aneurysm interventions," in *Biomedical Engineering and Informatics (BMEI), 2011 4th International Conference on*, vol. 1, pp. 550 –555, oct. 2011.

[40] H. Livyatan, Z. Yaniv, and L. Joskowicz, "Gradient-based 2-d/3-d rigid registration of fluoroscopic x-ray to ct," *Medical Imaging, IEEE Transactions on*, vol. 22, pp. 1395 –1406, nov. 2003.

[41] S. Ganapathy, "Decomposition of transformation matrices for robot vision," in *Robotics and Automation. Proceedings. 1984 IEEE International Conference on*, vol. 1, pp. 130 – 139, mar 1984.

[42] P. J. Besl and N. D. McKay, "A method for registration of 3-d shapes," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 14, pp. 239–256, February 1992.

[43] D. Ruijters, B. ter Haar Romeny, and P. Suetens, "Vesselness-based 2d–3d registration of the coronary arteries," *International Journal of Computer Assisted Radiology and Surgery*, vol. 4, pp. 391–397, 2009.

[44] O. Skrinjar, "Point-based registration with known correspondence: Closed form optimal solutions and properties," in *Biomedical Image Registration* (J. Pluim, B. Likar, and F. Gerritsen, eds.), vol. 4057 of *Lecture Notes in Computer Science*, pp. 315–321, Springer Berlin / Heidelberg, 2006.

[45] D. Simard, "Étude du recalage 3d-2d basé sur des repères externes dans une suite xmr," Master's thesis, 2005.

[46] K. S. Rhode, D. L. G. Hill, P. J. Edwards, J. Hipwell, D. Rueckert, G. Sanchez-Ortiz, S. Hegde, V. Rahunathan, and R. Razavi, "Registration and tracking to integrate x-ray and mr images in an xmr facility," *Medical Imaging, IEEE Transactions on*, vol. 22, no. 11, pp. 1369–1378, 2003.

[47] H. Lamecker, T. Wenckebach, and H.-C. Hege, "Atlas-based 3d-shape reconstruction from x-ray images," in *Pattern Recognition, 2006. ICPR 2006. 18th International Conference on*, vol. 1, pp. 371 –374, 0-0 2006.

[48] O. Sadowsky, G. Chintalapani, and R. H. Taylor, "Deformable 2d-3d registration of the pelvis with a limited field of view, using shape statistics," in *Proceedings of the 10th international conference on Medical image computing and computer-assisted intervention*, MICCAI'07, (Berlin, Heidelberg), pp. 519–526, Springer-Verlag, 2007.

[49] J. Yao and R. Taylor, "Assessing accuracy factors in deformable 2d/3d medical image registration using a statistical pelvis model," in *Computer Vision, 2003. Proceedings. Ninth IEEE International Conference on*, pp. 1329 –1334 vol.2, oct. 2003.

[50] G. Penney, P. Edwards, J. Hipwell, M. Slomczykowski, I. Revie, and D. Hawkes, "Postoperative calculation of acetabular cup position using 2-d ndash;3-d registration," *Biomedical Engineering, IEEE Transactions on*, vol. 54, pp. 1342 –1348, july 2007.

[51] M. Fleute and S. Lavallée, "Nonrigid 3-d/2-d registration of images using statistical models," in *Medical Image Computing and Computer-Assisted Intervention – MICCAI'99* (C. Taylor and A. Colchester, eds.), vol. 1679 of *Lecture Notes in Computer Science*, pp. 138–147, Springer Berlin / Heidelberg, 1999.

[52] T. Tang and R. Ellis, "2d/3d deformable registration using a hybrid atlas," in *Medical Image Computing and Computer-Assisted Intervention – MICCAI 2005* (J. Duncan and G. Gerig, eds.), vol. 3750 of *Lecture Notes in Computer Science*, pp. 223–230, Springer Berlin / Heidelberg, 2005.

[53] G. Zheng, M. Ballester, M. Styner, and L.-P. Nolte, "Reconstruction of patient-specific 3d bone surface from 2d calibrated fluoroscopic images and point distribution model," in *Medical Image Computing and Computer-Assisted Intervention – MICCAI 2006* (R. Larsen, M. Nielsen, and J. Sporring, eds.), vol. 4190 of *Lecture Notes in Computer Science*, pp. 25–32, Springer Berlin / Heidelberg, 2006.

[54] B. Jaramaz and K. Eckman, "2d/3d registration for measurement of implant alignment after total hip replacement," in *Medical Image Computing and Computer-Assisted Intervention – MICCAI 2006* (R. Larsen, M. Nielsen, and J. Sporring, eds.), vol. 4191 of *Lecture Notes in Computer Science*, pp. 653–661, Springer Berlin / Heidelberg, 2006.

[55] T. Yamazaki, T. Watanabe, Y. Nakajima, K. Sugamoto, T. Tomita, H. Yoshikawa, and S. Tamura, "Improvement of depth position in 2-d/3-d registration of knee implants using single-plane fluoroscopy," *Medical Imaging, IEEE Transactions on*, vol. 23, pp. 602 –612, may 2004.

[56] S. Benameur, M. Mignotte, S. Parent, H. Labelle, W. Skalli, and J. de Guise, "3d/2d registration and segmentation of scoliotic vertebrae using statistical models," *Computerized Medical Imaging and Graphics*, vol. 27, no. 5, pp. 321 – 337, 2003.

[57] J. Jomier, E. Bullitt, M. Van Horn, C. Pathak, and S. Aylward, "3d/2d model-to-image registration applied to tips surgery," in *Medical Image Computing and Computer-Assisted Intervention – MICCAI 2006* (R. Larsen, M. Nielsen, and J. Sporring, eds.), vol. 4191 of *Lecture Notes in Computer Science*, pp. 662–669, Springer Berlin / Heidelberg, 2006.

[58] M. Vermandel, N. Betrouni, G. Palos, J.-Y. Gauvrit, C. Vasseur, and J. Rousseau, "Registration, matching, and data fusion in 2d/3d medical imaging: Application to dsa and mra," in *Medical Image Computing and Computer-Assisted Intervention - MICCAI 2003* (R. Ellis and T. Peters, eds.), vol. 2878 of *Lecture Notes in Computer Science*, pp. 778–785, Springer Berlin / Heidelberg, 2003.

[59] D. Thivierge-Gaulin, C.-R. Chou, A. Kiraly, C. Chefd'Hotel, N. Strobel, and F. Cheriet, "3d-2d registration based on mesh-derived image bisection," in *Biomedical Image Registration* (B. Dawant, G. Christensen, J. Fitzpatrick, and D. Rueckert, eds.), vol. 7359 of *Lecture Notes in Computer Science*, pp. 70–78, Springer Berlin / Heidelberg, 2012.

[60] M. Woo, J. Neider, T. Davis, and D. Shreiner, *OpenGL Programming Guide: The Official Guide to Learning OpenGL, Version 1.2*. Boston, MA, USA: Addison-Wesley Longman Publishing Co., Inc., 3rd ed., 1999.

[61] T. Chan and L. Vese, "Active contours without edges," *Image Processing, IEEE Transactions on*, vol. 10, no. 2, pp. 266 –277, 2001.

[62] T. Akenine-Moller, T. Moller, and E. Haines, *Real-Time Rendering*. Natick, MA, USA: A. K. Peters, Ltd., 2nd ed., 2002.

[63] Kitware, Inc., *ParaView 3.14.1, A Parallel Visualization Application*, 2012.